

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:55:48 ; Search time 41 Seconds
(without alignments)
998.816 Million cell updates/sec

Title: US-09-813-453B-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNNTNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1335	100.0	258	AAU01243	B. subtilis novel
2	1335	100.0	258	AAU91149	Bacillus subtilis
3	1114	83.4	233	AAU91163	Pantothenate kinas
4	1046	78.4	258	AAU91172	Pantothenate kinas
5	1034	77.5	262	AAU91170	Pantothenate kinas
6	987	73.9	254	AAU91171	Pantothenate kinas
7	852	63.8	259	ABB47661	Listeria monocytog
8	794.5	59.5	256	AAU91175	Pantothenate kinas
9	756	56.6	255	AAU91154	Geobacter sulfurre

10	655	49.1	265	23	AAU91151	Streptomyces coeli
11	646.5	48.4	250	23	AAU91150	Clostridium acetob
12	577	43.2	260	23	AAU91173	Pantothenate kinas
13	539.5	40.4	256	23	ABP65945	Bifidobacterium lo
14	537	40.2	258	23	AAU91153	Rhodobacter capsul
15	515	38.6	219	23	AAU91176	Pantothenate kinas
16	494.5	37.0	272	23	AAU91152	Mycobacterium tube
17	493.5	37.0	272	22	AAG81225	Mycobacterium tube
18	428	32.1	262	23	AAU91155	Deinococcus radiop
19	405.5	30.4	246	23	AAU91156	Thermotoga maritim
20	338.5	25.4	212	23	AAU91177	Pantothenate kinas
21	327.5	24.5	273	23	AAU91157	Pantothenate kinas
22	319	23.9	257	23	AAU91174	Treponema pallidum
23	271.5	20.3	262	23	AAU91158	Pantothenate kinas
24	207	15.5	244	23	AAU91168	Borrelia burgdorfe
25	203	15.2	241	23	AAU91179	Pantothenate kinas
26	163	12.2	249	23	AAU91182	Pantothenate kinas
27	163	12.2	257	23	AAU91160	Synechocystis pant
28	154.5	11.6	229	23	AAU91159	Aquifex aeolicus p
29	154	11.5	249	23	AAU91178	Pantothenate kinas
30	151	11.3	455	20	AAV38617	Neisseria gonorrh
31	151	11.3	455	21	AAV74908	Neisseria gonorrh
32	151	11.3	460	23	AAU91167	Pantothenate kinas
33	151	11.3	592	20	AAV38618	Neisseria gonorrh
34	151	11.3	592	21	AAV74911	Neisseria gonorrh
35	151	11.3	592	24	ABP77010	N. gonorrhoeae ami
36	150	11.2	389	21	AAV74909	Neisseria meningit
37	150	11.2	455	21	AAV74910	Neisseria meningit
38	150	11.2	592	20	AAV38615	Neisseria meningit
39	150	11.2	592	20	AAV38616	Neisseria meningit
40	150	11.2	592	21	AAV74912	Neisseria meningit
41	150	11.2	592	21	AAV74913	Neisseria meningit
42	150	11.2	592	23	AAU91166	Pantothenate kinas
43	150	11.2	592	23	AAU91169	Pantothenate kinas
44	150	11.2	592	24	ABU06036	N. meningitidis va
45	134.5	10.1	242	23	AAU91180	Pantothenate kinas

ALIGNMENTS

RESULT 1
AAU01243
ID AAU01243 standard; Protein; 258 AA.
XX
AC AAU01243;
XX
DT 18-JUL-2001 (first entry)
XX
DE B. subtilis novel pantothenate kinase encoded by the gene coaX.
XX
KW Pantothenate kinase; coaX; pantothenate biosynthesis; vitamin B5;
XX nutritional supplement; panto-compound; pantoate.
OS Bacillus subtilis.
XX
PN WO200121772-A2.
XX
PD 29-MAR-2001.
XX
PF 21-SEP-2000; 2000WO-US25993.
XX
PR 21-SEP-1999; 99US-0400494.
PR 07-JUN-2000; 2000US-0210072.
PR 28-JUL-2000; 2000US-0221836.
PR 24-AUG-2000; 2000US-0227860.
XX (OMNI-) OMNIGENE BIOPRODUCTS.
PI Yocum RR, Patterson TA, Hermann T, Pero JG;
XX WPI; 2001-218644/22.
DR N-PSDB; AAS00984.

XX New recombinant microorganism which overexpress a Bacillus subtilis
PT pantothenate biosynthetic enzyme, useful for the high yield production
PT of panto-compounds such as pantothenate and pantoate -
XX
PS Example 14; Fig 23; 292pp; English.
XX
CC The sequence represents a novel B. subtilis pantothenate kinase (encoded
CC by gene coaX), an enzyme of the pantothenate biosynthetic pathway.
CC Pantothenate, also known as vitamin B5, is used as a nutritional
CC supplement in mammals and humans. The invention concerns methods of
CC producing recombinant microorganisms overexpressing at least one Bacillus
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods
CC of producing them are useful for producing a panto-compound such as
CC pantothenate or pantoate, which is a nutritional requirement for
CC livestock and humans. The methods are also useful for the identification
CC of pantothenate kinase modulators. Panto-compounds are produced at a
CC significantly higher yield than prior art methods and can be produced
CC independent of the need to feed precursors which decreases expense.
XX
SQ Sequence 258 AA;

Query Match 100.0%; Score 1335; DB 22; Length 258;
Best Local Similarity 100.0%; Pred. No. 7.7e-142;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60
Db 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60
Qy 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKEVGADRIVNAVAIHL 120
Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKEVGADRIVNAVAIHL 120
Qy 121 YGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 121 YGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
Qy 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Qy 241 LTLKGLELIYERNRVGSV 258
Db 241 LTLKGLELIYERNRVGSV 258

RESULT 2
AAU91149
ID AAU91149 standard; Protein; 258 AA.
XX
AC AAU91149;
XX
DT 05-JUN-2002 (first entry)
XX
DE Bacillus subtilis pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus subtilis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX

PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54168.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 67-68; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 258 AA;

Query Match 100.0%; Score 1335; DB 23; Length 258;
Best Local Similarity 100.0%; Pred. No. 7.7e-142;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60
Db 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60
Qy 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKEVGADRIVNAVAIHL 120
Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKEVGADRIVNAVAIHL 120
Qy 121 YGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 121 YGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
Qy 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Qy 241 LTLKGLELIYERNRVGSV 258
Db 241 LTLKGLELIYERNRVGSV 258

RESULT 3
AAU91163
ID AAU91163 standard; Protein; 233 AA.
XX
AC AAU91163;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #1.
DE
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus subtilis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX

PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX N-PSDB; ABK54169.
DR
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
PS
XX Disclosure; Page 81-82; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 233 AA;

Query Match 83.4%; Score 1114; DB 23; Length 233;
Best Local Similarity 99.5%; Pred. No. 5.7e-117;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLVIDVGNNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MLLVIDVGNNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120

QY 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIETRPDNI 180

QY 181 GKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQD 214
Db 181 GKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQD 214

RESULT 4
AAU911172
ID AAU911172 standard; Protein; 258 AA.
XX
AC AAU911172;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #10.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus stearothermophilus.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX

PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX N-PSDB; ABK54193.
DR
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
PS
XX Claim 10; Page 101-102; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 258 AA;

Query Match 78.4%; Score 1046; DB 23; Length 258;
Best Local Similarity 78.3%; Pred. No. 3.1e-109;
Matches 198; Conservative 32; Mismatches 23; Indels 0; Gaps 0;

QY 1 LLLVIDVGNNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MIFVLDVGNNTNTVLGVYDGDLELKHWRIETSRKTEDEYGMIMKALLNHVGLQFSDIRGI 60

QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120

QY 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGSPLIIVDFGTATTCYINEHKQYMGGAIAPGIMISTEALFARAACKLPRIETARPDII 180

QY 181 GKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKNTVSAMQAGILYGVYGVQVEGIVSRMKAKSKIIPPKVIATGGLAPLIASESDIIDVVDPF 240

QY 241 LTLKGLLEIYERN 253
Db 241 LTLTGLKLLYEKN 253

RESULT 5
AAU911170
ID AAU911170 standard; Protein; 262 AA.
XX
AC AAU911170;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #8.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus anthracis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX

XX	AAU911154;	ID	AAU911151 standard; Protein; 265 AA.
AC		XX	
XX		AC	AAU911151;
DT	05-JUN-2002 (first entry)	XX	
XX		DT	05-JUN-2002 (first entry)
DE	Geobacter sulfurreducens pantothenate kinase Coax.	XX	
XX		DE	Streptomyces coelicolor pantothenate kinase Coax.
XX	Pantothenate kinase; Coax; antibiotic; antimicrobial;	XX	
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.	KW	Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX		KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.
OS	Geobacter sulfurreducens.	XX	
XX		XX	Streptomyces coelicolor.
XX	WO200216601-A2.	OS	
PN		XX	WO200216601-A2.
XX		PN	
PD	28-FEB-2002.	XX	28-FEB-2002.
XX		PD	
XX	24-AUG-2001; 2001WO-US26531.	XX	
PF		PF	24-AUG-2001; 2001WO-US26531.
XX		XX	
PR	24-AUG-2000; 2000US-227860P.	XX	
XX		PR	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.	PR	20-MAR-2001; 2001US-0813453.
XX		XX	
XX	(OMNI-) OMNIGENE BIOPRODUCTS INC.	XX	
PA		PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX		XX	
PI	Yocum RR, Patterson TA;	PI	Yocum RR, Patterson TA;
XX		XX	
DR	WPI; 2002-269358/31.	XX	
XX		DR	WPI; 2002-269358/31.
XX	Identifying potential antibiotic or antimicrobial agent, comprises	XX	
PT	contacting composition comprising pantothenate kinase (Coax) protein	PT	Identifying potential antibiotic or antimicrobial agent, comprises
PT	with test compound and identifying inhibitor of the Coax protein -	PT	contacting composition comprising pantothenate kinase (Coax) protein
XX		PT	with test compound and identifying inhibitor of the Coax protein -
PS	Claim 10; Page 72-73; 128pp; English.	XX	
XX		PS	Claim 10; Page 69-70; 128pp; English.
CC	The invention describes assays for identifying a (potential) antibiotic	XX	
CC	comprising contacting an assay composition comprising a pantothenate	CC	The invention describes assays for identifying a (potential) antibiotic
CC	kinase (Coax) protein with a test compound, and determining the ability	CC	comprising contacting an assay composition comprising a pantothenate
CC	of the test compound to inhibit the activity of the Coax protein, an	CC	kinase (Coax) protein with a test compound, and determining the ability
CC	essential enzyme for the production of coenzyme A. Coax protein is a	CC	of the test compound to inhibit the activity of the Coax protein, an
CC	valuable target for identifying bactericidal compounds. Coax modulating	CC	essential enzyme for the production of coenzyme A. Coax protein is a
CC	agents can be used in an infectious animal model to determine the	CC	valuable target for identifying bactericidal compounds. Coax modulating
CC	efficacy, toxicity, or side effects of treatment with such an agent. This	CC	agents can be used in an infectious animal model to determine the
CC	is the amino acid sequence of a pantothenate kinase (Coax) protein	CC	efficacy, toxicity, or side effects of treatment with such an agent. This
CC	described in the invention.	CC	is the amino acid sequence of a pantothenate kinase (Coax) protein
XX		CC	described in the invention.
SQ	Sequence 255 AA;	XX	
		SQ	Sequence 265 AA;
<div>Query Match 56.6%; Score 756; DB 23; Length 255; Best Local Similarity 57.5%; Pred. No. 1.6e-76; Matches 146; Conservative 46; Mismatches 62; Indels 0; Gaps 0;</div>			
Qy	1 LLLVIDVGNNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60	Qy	1 LLLVIDVGNNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ---- 56
Db	1 MLLVIDVGNNTIVLGIYDGERLVRDWRVSTDKARTTDEYGILINELFRLAGLDQIRAV 60	Db	1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGHPLLGLDELGDG 60
Qy	61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKKEVGADRIVNAVAIHL 120	Qy	57 IDGIILISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNICYDNPKKEVGADRIVNAV 115
Db	61 IISVVPPILTGVLRLSLGYFGMRPLVVGPGIKTGMPIQYDNPREVGADRIVNAVAGYEK 120	Db	61 IDGIAICATVPSVLHELREVTTRYGDPVPAVLVEPGVKTGVPILTDPKKEVGADRIINAV 120
Qy	121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPIGITISTEALYSRAAKLPRIETRPDNI 180	Qy	116 AAHLHYGNPLIVDFGTATTCYIDENKQYMGGAIAPIGITISTEALYSRAAKLPRIETR 175
Db	121 YRTSLIIVDFGTATTFDYVNRKGEYCGGAIAPLGVISTEALFQRASKLPVDDIIRPSAI 180	Db	121 AAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPIGIEISVEALGVKGAQLRKIEVAR 180
Qy	181 GKNTVSAMQSGILFGYVGQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240	Qy	176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKQQA---KQDLKVIATGGLAPLIANESD 232
Db	181 ARNTVSNMQAGIYYGYVGLVDEIVTRMKAESKDAPRVIAATGGLASLIAPESKTIEAVEEY 240	Db	181 PRSVIGKNTVEAMQSGIVYGFAGQVDGVNRMARELADDDVTVIATGGLAPMVLGESS 240
Qy	241 LTLKGLELIYERNR 254	Qy	233 CIDIVDPFPLTLKGLELIYERN 253
Db	241 LTLEGLRIYERNR 254	Db	241 VIDEHEPWLTLMGLRLVYERN 261
<div>RESULT 10 AAU911151</div>			
RESULT 11			

AAU911150
ID AAU911150 standard; Protein; 250 AA.
XX AC AAU911150;
XX DT 05-JUN-2002 (first entry)
XX DE Clostridium acetobutylicum pantothenate kinase CoaX.
XX KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Clostridium acetobutylicum.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX DR WPI; 2002-269358/31.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX PS Claim 10; Page 68-69; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
SQ Sequence 250 AA;
Query Match 48.4%; Score 646.5; DB 23; Length 250;
Best Local Similarity 53.9%; Pred. No. 3.6e-64;
Matches 125; Conservative 46; Mismatches 56; Indels 5; Gaps 1;
QY 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 18 VILVLDVGNTNVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGV 77
QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAHIHL 120
Db 78 IISVVPPIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVAHEI 137
QY 121 YGNPLIVDFGTATTYCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNI 180
Db 138 YKRSIIIDFGTATTFCVRENGDYLGAICPGIKVSSEALFEKAALPRVELIKPAYAI 197
QY 181 GKNTVSAMQSGILFGYVGVEGIVKRMKWAQKQDLK-----VIATGGLAPLI 227
Db 198 CKNTISSIQSIGIVRYLRQVKYLFKLEKLENLPDGRTRTSLVLATGGLAKLI 249
RESULT 12
AAU911173
ID AAU911173 standard; Protein; 260 AA.
XX

AAU911173;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (CoaX) #11.
XX KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Caulobacter crescentus.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX DR WPI; 2002-269358/31.
XX DR N-PSDB; ABK54194.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX PS Claim 10; Page 102-103; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
SQ Sequence 260 AA;
Query Match 43.2%; Score 577; DB 23; Length 260;
Best Local Similarity 44.4%; Pred. No. 2.6e-56;
Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;
QY 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MLLAIEQGNTNTMFAIHGASWVAQWRSATSTRTADEYVWLSQLLSMQGLGFRDAIDAV 60
QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAHIHL 120
Db 61 IISVVPPQSIFNLRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120
QY 121 YGNPLIVDFGTATTYCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRP--DN 178
Db 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180
QY 179 IIGKNTVSAMQSGILFGYVGVEGIVKRMKWAQKQDLKVIATGGLAPLIANESDCIDIVD 238
Db 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPTMTVIATGGVASLFEGATDSIDHFD 240
QY 239 PFLTLLKGLLEIYERNRV 255
Db 241 SDLTIRGLLEIYERNNTI 257
RESULT 13
ABP65945

Db 60 IISSTAPRVVFNLRVLCNRYFDRCRPYVVGKPGCELPVAPRVDPGTTVGPDRLVNTVAGYD 119
Qy 120 LYGNPLIVVDFGTATTVCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 179
Db 120 RHGGDLIVVDFGTATTVDVAPDGAAYIGGVIAPGVNLSEALHMAAALPHVDVTKPGV 179
Qy 180 IGKNTVSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDP 239
Db 180 IGTNTVACIQSGVYWGVIYGLVEGIVRQIRMERDRPMKVIAATGGLASLFDLGLFDLKVED 239
Qy 240 FLTLKGLELIYERNR 254
Db 240 DLTMHGLRLIFDYNK 254

RESULT 15
AAU91176
ID AAU91176 standard; Protein; 219 AA.
XX
AC AAU91176;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #14.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Dehalococcoides ethenogenes.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54197.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
PS Claim 10; Page 106-107; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.

SQ Sequence 219 AA;
Query Match 38.6%; Score 515; DB 23; Length 219;
Best Local Similarity 45.8%; Pred. No. 2e-49;
Matches 97; Conservative 45; Mismatches 70; Indels 0; Gaps 0;
Qy 2 LLVIDVGNNTVLGVYHDKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGII 61
Db 5 LVAVDIGNTSVNIGIFEKGLANWHLGSAQRMADYASLLGLLQHAGIHPEELNRVI 64

Qy 62 ISSVVPPIMPFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAAIHLY 121
Db 65 MCSVVPPLTTTPEEVFKSYFKAAPLVVVGAGIKSGVKVRMDNPREVGADRIVNAARVLY 124
Qy 122 GNPLIVVDFGTATTVCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 181
Db 125 PGACIIIVDMGTATTFDTLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPAK 184
Qy 182 KNTVSAMQSGILFGYVQVEGIVKRMKWOAKQ 213
Db 185 SNTVSAMQSGIYFGYIGLVEELVRRRIQTCLGQ 216

Search completed: December 18, 2003, 13:01:37
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:54:22 ; Search time 31 Seconds
(without alignments)
1554.314 Million cell updates/sec

Title: US-09-813-453B-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNNTVLGVYHDG.....PFLTLKGLELIYERNRVCVSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1335	100.0	258	10	US-09-813-453A-2 Sequence 2, Appli
2	1114	83.4	233	10	US-09-813-453A-17 Sequence 17, Appli
3	1046	78.4	258	10	US-09-813-453A-49 Sequence 49, Appli
4	1034	77.5	262	10	US-09-813-453A-45 Sequence 45, Appli
5	987	73.9	254	10	US-09-813-453A-47 Sequence 47, Appli
6	794.5	59.5	256	10	US-09-813-453A-55 Sequence 55, Appli
7	756	56.6	255	10	US-09-813-453A-7 Sequence 7, Appli
8	655	49.1	265	10	US-09-813-453A-4 Sequence 4, Appli
9	655	49.1	265	15	US-10-156-761-12224 Sequence 12224, A
10	646.5	48.4	250	10	US-09-813-453A-3 Sequence 3, Appli
11	577	43.2	260	10	US-09-813-453A-51 Sequence 51, Appli
12	537	40.2	258	10	US-09-813-453A-6 Sequence 6, Appli
13	515	38.6	219	10	US-09-813-453A-57 Sequence 57, Appli
14	494.5	37.0	272	10	US-09-813-453A-5 Sequence 5, Appli
15	493.5	37.0	272	10	US-09-712-363-276 Sequence 276, App

16	428	32.1	262	10	US-09-813-453A-8	Sequence 8, Appli
17	405.5	30.4	246	10	US-09-813-453A-9	Sequence 9, Appli
18	338.5	25.4	212	10	US-09-813-453A-59	Sequence 59, Appli
19	327.5	24.5	273	10	US-09-813-453A-10	Sequence 10, Appli
20	319	23.9	257	10	US-09-813-453A-53	Sequence 53, Appli
21	271.5	20.3	262	10	US-09-813-453A-11	Sequence 11, Appli
22	207	15.5	244	10	US-09-813-453A-41	Sequence 41, Appli
23	203	15.2	241	10	US-09-813-453A-63	Sequence 63, Appli
24	163	12.2	249	10	US-09-813-453A-70	Sequence 70, Appli
25	163	12.2	257	10	US-09-813-453A-13	Sequence 13, Appli
26	154.5	11.6	229	10	US-09-813-453A-12	Sequence 12, Appli
27	154	11.5	249	10	US-09-813-453A-61	Sequence 61, Appli
28	151	11.3	460	10	US-09-813-453A-39	Sequence 39, Appli
29	150	11.2	592	10	US-09-813-453A-22	Sequence 22, Appli
30	150	11.2	592	10	US-09-813-453A-43	Sequence 43, Appli
31	150	11.2	592	12	US-10-320-800-48	Sequence 48, Appli
32	134.5	10.1	242	10	US-09-813-453A-65	Sequence 65, Appli
33	133	10.0	248	10	US-09-813-453A-20	Sequence 20, Appli
34	129.5	9.7	267	10	US-09-813-453A-15	Sequence 15, Appli
35	109.5	8.2	223	10	US-09-895-913A-74	Sequence 74, Appli
36	109.5	8.2	223	10	US-09-813-453A-14	Sequence 14, Appli
37	109.5	8.2	223	10	US-09-813-453A-67	Sequence 67, Appli
38	109	8.2	209	10	US-09-813-453A-21	Sequence 21, Appli
39	80	6.0	449	9	US-09-815-242-5474	Sequence 5474, Ap
40	80	6.0	449	9	US-09-815-242-12348	Sequence 12348, A
41	80	6.0	449	9	US-09-815-242-12793	Sequence 12793, A
42	79	5.9	636	9	US-09-205-658-160	Sequence 160, App
43	79	5.9	636	12	US-09-963-693-160	Sequence 160, App
44	79	5.9	923	15	US-10-262-538-6	Sequence 6, Appli
45	78.5	5.9	337	15	US-10-075-846-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1

US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn.Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 100.0%; Score 1335; DB 10; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.6e-144;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	LLLVIDVGNNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI	60
Db	1	LLLVIDVGNNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI	60
QY	61	IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL	120
Db	61	IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL	120
QY	121	YGNPLIVDPGTATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETRPDNI	180

Db 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
QY 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
QY 241 LTLKGLELIYERNRGSV 258
Db 241 LTLKGLELIYERNRGSV 258

RESULT 2
US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-17

Query Match 83.4%; Score 1114; DB 10; Length 233;
Best Local Similarity 99.5%; Pred. No. 3.2e-119;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKKEVGADRIVNAVAIHL 120
Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKKEVGADRIVNAVAIHL 120
QY 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
QY 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214
Db 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214

RESULT 3
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49
Query Match 78.4%; Score 1046; DB 10; Length 258;
Best Local Similarity 78.3%; Pred. No. 2.3e-111;
Matches 198; Conservative 32; Mismatches 23; Indels 0; Gaps 0;
QY 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MIFVLDVGNTNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMIMKALLNHVGLQFSDIRGI 60
QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKKEVGADRIVNAVAIHL 120
Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKKEVGADRIVNAVAIHL 120
QY 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAPRIEITARPDDII 180
QY 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKNTVSAMQAGILYGVYQVEGIVSRMKAKSKIPTKVIATGGLAPLIASESDIIDVVDPF 240
QY 241 LTLKGLELIYERN 253
Db 241 LTLTGLKLLYEKN 253

RESULT 4
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453A-45

Query Match 77.5%; Score 1034; DB 10; Length 262;
Best Local Similarity 75.1%; Pred. No. 5.7e-110;
Matches 190; Conservative 34; Mismatches 29; Indels 0; Gaps 0;
QY 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MIFVLDVGNTNTVLGVFEELRQHRMETDRHKTEDEYGMILVKQLLEHGLSFEDVKGI 60
QY 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNICYDNPKKEVGADRIVNAVAIHL 120
Db 61 IVSSVPPIMFALERMCEKYFKIKPLVLVVGPGIKTGLNICYENPREVGADRIVNAVAIHL 120
QY 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 121 YGSPLIIVDFGTATTTCYINEEKHYMGGVIPTGIMISAEALYSRAAKLPRIEITKPSVV 180
QY 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240

Db 181 GKNVTSAMQSGILYGVYQVEGIVKRMKEAKQEPKVATGGLAKLISEESNVIDVDPF 240
QY 241 LTLKGLELIYERN 253
Db 241 LTLKGLYMLYERN 253

RESULT 5
US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match 73.9%; Score 987; DB 10; Length 254;
Best Local Similarity 72.3%; Pred. No. 1.3e-104;
Matches 183; Conservative 33; Mismatches 37; Indels 0; Gaps 0;
QY 1 LLLVIDVGNNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MLLVIDVGNNTVLGVYQDETIVHHRLATSRQKTEDEYAMTVRSLEFDHAGLQFQDIDGI 60
QY 61 IISSVVPPIFMALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db 61 VISSVPPMFMFSLQEMCKKYFHVTPMIIIGPGIKTGLNPKYDNPKEVGADRIVNAVAIHL 120
QY 121 YGNPLIVDFGTATTYCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGYPATVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIETIAKPKQV 180
QY 181 GKNVTSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GTNTIDSMQSGIFYGVYQVDSQVGVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSF 240
QY 241 LTLKGLELIYERN 253
Db 241 LTLKGLQLIYKKN 253

RESULT 6
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55

Query Match 59.5%; Score 794.5; DB 10; Length 256;
Best Local Similarity 60.8%; Pred. No. 1.7e-82;
Matches 155; Conservative 40; Mismatches 59; Indels 1; Gaps 1;
QY 1 LLLVIDVGNNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MLLVFDVGNNTMVLGIYKGDKLNVYWRIKTDREKTSDEYGILISNLFYDYNVNISDIDDV 60
QY 61 IISSVVPPIFMALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db 61 IISSVVPNVHMSLENFCIKYCKKQPLIVGPGIKTGLNPKYDNPKEVGADRIVNAVAGIEK 120
QY 121 YGNPLIVDFGTATTYCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGAPSILVDFGTATTFCAISEKGEYLGGTIAPGIKISSEALFQSASKLPRLVELAKPGMTI 180
QY 181 GKNVTSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDP 239
Db 181 CKSTVSAMQSGIIVGYVGLVDKIISIMKKELNCCDDVKVIATGGLAKLIASETKSIDYVDG 240
QY 240 FLTLKGLELIYERN 254
Db 241 FLTLEGLRIIYEKNQ 255

RESULT 7
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 56.6%; Score 756; DB 10; Length 255;
Best Local Similarity 57.5%; Pred. No. 4.3e-78;
Matches 146; Conservative 46; Mismatches 62; Indels 0; Gaps 0;
QY 1 LLLVIDVGNNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MLLVIDVGNNTIVLGIYDGERLVRVSTDKARTTDEYGILINELFRLAGLGLDQIRAV 60
QY 61 IISSVVPPIFMALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db 61 IISSVVPPLTGVLERLSLGYFGMRPLVVGPGIKTGMPIQYDNPKEVGADRIVNAVAGYEK 120
QY 121 YGNPLIVDFGTATTYCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNI 180
Db 121 YRTSLIIVDFGTATTFDYNRKEGCGGAIAPGLVISTEALFORASKLPRLVDIIRPSAI 180

QY 182 KNTVSAMQSGILFGYVGVQVEGIVKRMKQAKQ 213
Db 185 SNTVSAMQSGIYFGYIGLVEELVRIQTELQ 216

RESULT 14

US-09-813-453A-5
; Sequence 5, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

Query Match 37.0%; Score 494.5; DB 10; Length 272;
Best Local Similarity 38.8%; Pred. No. 4.8e-48;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

QY 1 LLLVIDVGNNTNTVLGVY---HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
Db 1 MLLAIDVRNTHTVVGLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDS---ER 57
QY 57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
Db 58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIIEPGVRTGIPLLVDNPKEVGADRIVNCL 117
QY 116 AAHLHYGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITR 175
Db 118 AAYDRFRKAAIVVDFGSSICVDVVSAGKEFLGGAIAPGVQVSSDAAAARSAAALRRVELAR 177
QY 176 PDNIIGKNTVSAMQSGILFGYVGVQVEGIVKRMK-----WQAKQDLKVIATGGLAPLIANE 230
Db 178 PRSVVGKNTVECMQAGAVFGFAGLVGDLVGRIREDSVGSFSDVDHVAIVATGHTAPLLLPE 237
QY 231 SDCIDIVDPFLLTKGLELIYERN 253
Db 238 LHTVDHYDQHLTLQGLRLVFERN 260

RESULT 15

US-09-712-363-276
; Sequence 276, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844

; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 276
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-276

Query Match 37.0%; Score 493.5; DB 10; Length 272;
Best Local Similarity 38.8%; Pred. No. 6.2e-48;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

QY 1 LLLVIDVGNNTNTVLGVY---HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
Db 1 VLLAIDVRNTHTVVGLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDS---ER 57
QY 57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
Db 58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIIEPGVRTGIPLLVDNPKEVGADRIVNCL 117
QY 116 AAHLHYGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITR 175
Db 118 AAYDRFRKAAIVVDFGSSICVDVVSAGKEFLGGAIAPGVQVSSDAAAARSAAALRRVELAR 177
QY 176 PDNIIGKNTVSAMQSGILFGYVGVQVEGIVKRMK-----WQAKQDLKVIATGGLAPLIANE 230
Db 178 PRSVVGKNTVECMQAGAVFGFAGLVGDLVGRIREDSVGSFSDVDHVAIVATGHTAPLLLPE 237
QY 231 SDCIDIVDPFLLTKGLELIYERN 253
Db 238 LHTVDHYDQHLTLQGLRLVFERN 260

Search completed: December 18, 2003, 13:00:43
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: December 18, 2003, 13:00:04 ; Search time 21 Seconds
(without alignments)
1181.501 Million cell updates/sec

Title: US-09-813-453B-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNTNTVLGVYHDG.....PFLTLKGLELIYERNVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1114	83.4	233	2 S66100	conserved hypothet
2	987	73.9	254	2 F83660	hypothetical prote
3	852	63.8	259	2 AF1102	conserved hypothet
4	847	63.4	259	2 AF1464	conserved hypothet
5	775	58.1	273	2 E97293	probable transcrip
6	655	49.1	265	2 T36391	hypothetical prote
7	577	43.2	261	2 B87489	transcription acti
8	519.5	38.9	274	2 H86937	conserved hypothet
9	494.5	37.0	272	2 A70955	hypothetical prote
10	428	32.1	262	2 E75516	conserved hypothet
11	405.5	30.4	246	2 D72320	conserved hypothet
12	327.5	24.5	273	2 D71326	conserved hypothet
13	271.5	20.3	262	2 F70165	conserved hypothet
14	170	12.7	276	2 AI2292	hypothetical prote
15	163	12.2	257	2 S75559	hypothetical prote
16	154.5	11.6	229	2 E70465	hypothetical prote
17	150	11.2	592	2 B81009	BirA protein/Bvg a
18	150	11.2	592	2 H82031	probable biotin-la
19	134.5	10.1	242	2 A82637	conserved hypothet
20	133	10.0	248	2 H83111	hypothetical prote
21	129.5	9.7	267	2 I40327	baf protein - Bord
22	117	8.8	224	2 A99571	conserved hypothet
23	116.5	8.7	223	2 G71887	hypothetical prote
24	109.5	8.2	223	2 F64627	hypothetical prote
25	109	8.2	209	2 H81382	hypothetical prote
26	104.5	7.8	597	2 B69251	probable electron
27	95	7.1	467	2 AE3142	hypothetical prote
28	95	7.1	520	2 G98145	probable aminotran
29	93	7.0	1014	2 H71602	protein with DnaJ

30	91.5	6.9	828	2 G82583	bifunctional aspar
31	90.5	6.8	4452	1 YGBSG2	gramicidin S synth
32	88	6.6	256	2 H83002	imidazoleglycerol-
33	87.5	6.6	287	2 D90540	glucokinase (gluco
34	87	6.5	336	2 AE0625	dihydroorotate deh
35	86.5	6.5	205	2 AC0612	anaerobic dimethyl
36	86.5	6.5	205	2 AG0680	probable dimethyl
37	86.5	6.5	243	2 D64300	(R)-2-hydroxygluta
38	86	6.4	336	1 S13824	dihydroorotate oxi
39	85	6.4	336	1 DEECD0	dihydroorotate oxi
40	85	6.4	336	2 E90757	dihydro-orotate de
41	85	6.4	336	2 C85621	dihydro-orotate de
42	84.5	6.3	317	2 E96792	probable cinnamoyl
43	84	6.3	443	1 JCS298	protein kinase (EC
44	83.5	6.3	205	2 B85615	hypothetical prote
45	83.5	6.3	205	2 D90751	hypothetical prote

ALIGNMENTS

RESULT 1

S66100 conserved hypothetical protein yacB - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000

C;Accession: S66100; E69740

R;Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994

A;Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom

A;Reference number: S65967; MUID:96051385; PMID:7584024

A;Accession: S66100

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-233 <OGA>

A;Cross-references: EMBL:D26185; NID:G467326; PIDN:BAA05305.1; PID:d1005847; PID:g467459

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: E69740

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-233 <KUN>

A;Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:e1182003;

A;Experimental source: strain 168

C;Genetics:

A;Gene: yacB

A;Start codon: TTG

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 83.4%; Score 1114; DB 2; Length 233;

Best Local Similarity 99.5%; Pred. No. 3.1e-88;

Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
:|||||

Db 1 MLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
:|||||

Qy 61 IISVVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120

Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy 121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 180

Db 121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 180

Qy 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214

Db 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214

RESULT 2

F83660

hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C;Accession: F83660

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: F83660

A;Status: preliminary

A:Molecule type: DNA

A;Residues: 1-254 <STO>

A;Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BAB03805.1; GSPDB:GN00

A;Experimental source: strain C-125

C;Genetics:

A:Gene: BH0086

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 73.9%; Score 987; DB 2; Length 254;

Best Local Similarity 72.3%; Pred. No. 2.8e-77;

Matches 183; Conservative 33; Mismatches 37; Indels 0; Gaps 0;

Qy 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Db 1 MILVIDVGNTNTVLGVYQDETLLVHHRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGI 60

Qy 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Db 61 VISSVPPMFMFSLQMKCKYFHVTPMIGPIKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy 121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 180

Db 121 YGYPALIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQV 180

Qy 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240

Db 181 GTNTIDSMQSGIFYGVSVQVDGVVVRMKAQAESEPKVIATGGLAKLIGTETIDVIDSF 240

Qy 241 LTLKGLLELIYERN 253

Db 241 LTLKGLQLIYKKN 253

RESULT 3

AF1102

conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EGD-e

C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C;Accession: AF1102

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

Science 294, 849-852, 2001

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AF1102

A;Status: preliminary

A:Molecule type: DNA

A;Residues: 1-259 <GLA>

A;Cross-references: GB:NC_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lmo0221

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 63.8%; Score 852; DB 2; Length 259;

Best Local Similarity 63.4%; Pred. No. 1.1e-65;

Matches 161; Conservative 37; Mismatches 56; Indels 0; Gaps 0;

Qy 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Db 1 MILVIDVGNTNTCTGVVYEKQKLLKHWMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

Qy 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Db 61 IISVVPPIMHAMETMVCVRVFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

Qy 121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 180

Db 121 YGTPVIVVDFGTATTCFYIDESGVYQGGAIAPGIMISTEALYNRAAKLPVVDIAESSQII 180

Qy 181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240

Db 181 GKSTVSSMQAGIFYGVGQCEGIIAEMKKQSNASPVVATGGLARMITKSSAVDILDPPF 240

Qy 241 LTLKGLLELIYERNR 254

Db 241 LTLKGLLELLYRRNK 254

RESULT 4

AF1464

conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip11262)

C;Species: Listeria innocua

C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C;Accession: AF1464

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AF1464

A;Status: preliminary

A:Molecule type: DNA

A;Residues: 1-259 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178

A;Experimental source: strain Clip11262

C;Genetics:

A:Gene: lin0253

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 63.4%; Score 847; DB 2; Length 259;

Best Local Similarity 62.6%; Pred. No. 3e-65;

Matches 159; Conservative 39; Mismatches 56; Indels 0; Gaps 0;

Qy 1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Db 1 MILVIDVGNTNTCTGVVYEKQKLLRHWMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

Qy 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Db 61 IISVVPPIMHAMETMVCVRVFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

Qy 121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 180

Db 121 YGTPVIVVDFGTATTCFYIDEGVYQGGAIAPGIMISTEALYNRAAKLPVVDIAESSQII 180

QY 181 GKNTVSAMQSGILFGYVCQVEGIVKRMKWAQKDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKSTVASMQAGIFYGFIGQCEGIIAEMKKQSNTSPVVVATGGLARMITKSSAVIDLPF 240
QY 241 LTLKGLELIYERNR 254
Db 241 LTLKGLELLYRRNK 254

RESULT 5
E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: E97293
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: E97293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3200
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 58.1%; Score 775; DB 2; Length 273;
Best Local Similarity 56.4%; Pred. No. 4.8e-59;
Matches 146; Conservative 55; Mismatches 54; Indels 4; Gaps 1;

QY 1 LLLVIDVGNNTNTVLGVYHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 12 VILVDVGNNTNVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGV 71

QY 61 IISVVPPIMPALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120
Db 72 IISVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVAHEI 131

QY 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRPDNI 180
Db 132 YKRSIIIDFGTATTCFAVRENGDYLGAICPGIKVSSEALFEKAAKLPVRELKPAYAI 191

QY 181 GKNTVSAMQSGILFGYVCQVEGIVKRMKWAQKDLK---VIATGGLAPLIANESDCIDI 236
Db 192 CKNTISSIQSGIVGYGQVRYIVERMKEELOEKEKEPLVAVATGGLAKLISEAKNVDV 251

QY 237 VDPFLTTLKGLELIYERNRV 255
Db 252 INPFLTLEGLRIIYKNRV 270

RESULT 6
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T36391
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21573
A;Accession: T36391
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-265 <OLI>
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE94.31c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 49.1%; Score 655; DB 2; Length 265;
Best Local Similarity 51.0%; Pred. No. 9.4e-49;
Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3;

QY 1 LLLVIDVGNNTNTVLGVYHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQ---- 56
Db 1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRRTADELAVLLQGLMGHPLLGDELGDG 60

QY 57 IDGIIISSVVPPIMPALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
Db 61 IDGIAICATVPSVLHELREVTTRYGDPVPAVLVEPGVKTGPILTDHPKEVGADRIINAV 120

QY 116 AAIHLYGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITR 175
Db 121 AAVELYGGPAIVVDFGTATTTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKIEVAR 180

QY 176 PDNIIGKNTVSAMQSGILFGYVQVEGIVKRMKWAQ---KQDLKVIATGGLAPLIANESD 232
Db 181 PRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDPDDVTVIATGGLAPMVLGESS 240

QY 233 CIDIVDPFLTTLKGLELIYERN 253
Db 241 VIDEHEPWLTLMGLRLVYERN 261

RESULT 7
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
A;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1935
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 43.2%; Score 577; DB 2; Length 261;
Best Local Similarity 44.4%; Pred. No. 4.6e-42;
Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;

QY 1 LLLVIDVGNNTNTVLGVYHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 2 MLLAIEQGNNTNTMFAIHGASWAQWRSATSTRTADEYVVVWLSQLLSMQGLGFRAIDAV 61

QY 61 IISVVPPIMPALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120
Db 62 IISVVVPQSIFNLNLSRRYFNVEPLVIGENAKLIGIDVRIEKPSEAGADRLVNAIGAAMV 121

QY 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEITRP--DN 178
Db 122 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQRPAGNR 181

QY 179 IIGKNTVSAMQSGILFGYVQVEGIVKRMKWAQKDLKVIATGGLAPLIANESDCIDIVD 238
Db 182 IVGTDTVSAMQSGVFWGYISLIEGLVARIKARGEPMPTVIATGGVASLFEGATDSIDHFD 241

QY 239 PFLTTLKGLELIYERNRV 255
Db 242 SDLTIRGLLEIYRNTI 258

RESULT 8

H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H86937
R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A: Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A: Title: Massive gene decay in the leprosy bacillus.
A: Reference number: A86909; MUID: 21128732; PMID: 11234002
A: Accession: H86937
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-274 <STO>
A: Cross-references: GB:AL450380; NID: g13092576; PIDN: CAC29740.1; GSPDB: GN00147
C: Genetics:
A: Gene: ML0232
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 38.9%; Score 519.5; DB 2; Length 274;
Best Local Similarity 41.0%; Pred. No. 4.2e-37;
Matches 109; Conservative 53; Mismatches 89; Indels 15; Gaps 5;

Qy 1 LLLVIDVGNNTNTVLGVY---HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
Db 1 MLLAIDVRNTHTVVGLSGMKEHAKVVQWRIETSEVTADELALIDGLIGDDS---ER 57

Qy 57 IDGIISSVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKKEVGADRIVNAV 115
Db 58 LAGAAALSTVPSVLHEVRIMLDQWYPSVPHVLIIEPGVRTGIPLLVDNPKKEVGADRIVNCL 117

Qy 116 AAHLNGNPLIVDFGTATTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETR 175
Db 118 AAFHKFGQAIVVDFGSSICVDVVSAGFEFLGGAIAPGVQVSSDAAARSAAALRRVELAR 177

Qy 176 PDNIIGKNTVVSAMQSGILFGYVQVEGIVKRMKQAKQ---DL----KVIAATGGLAPLIA 228
Db 178 PRSVVGKNTVECMQAGVVFAGLVLDGLVGRMRQDVVEEFGDLGNRVAVVATGHTAPLLL 237

Qy 229 NESDCIDIVDPFLLTKGLELIYERNR 254
Db 238 PELHTVDHYDRHLTLHGLRLVFERN 263

RESULT 9
A70955
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70955
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: A70955
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-272 <COL>
A: Cross-references: GB:Z95557; GB:AL123456; NID: g3242276; PIDN: CAB08944.1; PID: g2113976
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv3600c
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 37.0%; Score 494.5; DB 2; Length 272;
Best Local Similarity 38.8%; Pred. No. 5.9e-35;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

Qy 1 LLLVIDVGNNTNTVLGVY---HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
Db 1 MLLAIDVRNTHTVVGLSGMKEHAKVVQWRIETSEVTADELALIDGLIGEDS---ER 57

Qy 57 IDGIISSVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKKEVGADRIVNAV 115
Db 58 LTGTAALSTVPSVLHEVRIMLDQWYPSVPHVLIIEPGVRTGIPLLVDNPKKEVGADRIVNCL 117

Qy 116 AAHLNGNPLIVDFGTATTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETR 175
Db 118 AAYDRFRKAAIVVDFGSSICVDVVSAGFEFLGGAIAPGVQVSSDAAARSAAALRRVELAR 177

Qy 176 PDNIIGKNTVVSAMQSGILFGYVQVEGIVKRMK-----WQAKQDLKVIATGGLAPLIANE 230
Db 178 PRSVVGKNTVECMQAGAVFGFAGLVLDGLVGRIREDEVSGFSVDHDAIVATGHTAPLLLE 237

Qy 231 SDCIDIVDPFLLTKGLELIYERN 253
Db 238 LHTVDHYDQHLTLQGLRLVFERN 260

RESULT 10
E75516
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C:Accession: E75516
R: White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Ma
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A: Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A: Reference number: A75250; MUID: 20036896; PMID: 10567266
A: Accession: E75516
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-262 <WHI>
A: Cross-references: GB:AE001905; GB:AE000513; NID: g6458144; PIDN: AAF10040.1; PID: g6458144
A: Experimental source: strain R1
C: Genetics:
A: Gene: DR0461
A: Map position: 1
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 32.1%; Score 428; DB 2; Length 262;
Best Local Similarity 37.4%; Pred. No. 2.9e-29;
Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;

Qy 2 LLLVIDVGNNTNTVLGVYH-DGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 6 LLAVDIGNTTTLVGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAGAPIPR--AA 63

Qy 61 IISVVVPI---MFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKKEVGADRIVNAVA 116
Db 64 VLSSVAPPVGENYALAKR---HFMIDAFAVSAENLPDVTVELDTPGSGVADRLCNLFG 119

Qy 117 AIHLYG--NPLIVDFGTATTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIET 174
Db 120 AEKYLGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRITLQ 179

Qy 175 RPDNIIGKNTVVSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCI 234
Db 180 APETAIGKNTVHALQSLGVLFGVAEMVDGLLRIRAEPLPGEAVAVATGGFSRTVQGICQEI 239

Qy 235 DIVDPFLLTKGL-ELIYERNV 255
Db 240 DYDETTLRLGLVELWASRSEV 261

RESULT 11
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: D72320
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing of *Streptomyces coelicolor* A72200; MUID:99287316; PMID:10360571
A;Accession: D72320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <ARN>
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g4981417
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0883
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.4%; Score 405.5; DB 2; Length 246;
Best Local Similarity 39.2%; Pred. No. 2.3e-27;
Matches 100; Conservative 42; Mismatches 88; Indels 25; Gaps 9;
QY 1 LLLVIDVGNNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSLFDH----SGLMFEQ 56
Db 1 MYLLVDVGNTHSVFSITEDCKTFRWRRLSTGVFQTEDE-----LFSLHPLLLGDAMRE 53
QY 57 IDGIISSVPPIMPFALERMCTKYFHIEPQIVGPGMKTGL--NIKYDNPKEVGADRIVN 113
Db 54 IKGIGVASVWPTQNTVIERFSQKYFHISP--IWVKAKNGCVKWNVK--NPSEVGADRIVAN 109
QY 114 AVAAIHLVGNPLIVVDFGTATTTCYIDENKQVMGGAIAPGITISTEALYSRAAKLPRIEI 173
Db 110 VVAFVKEYGKNGIIMDGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEV 168
QY 174 TRPDNIIGKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDC 233
Db 169 KPADFVVGKDTENIRLVGVSVALEGIIGRIK-EVYGDLPVLTGGQSKIV---KDM 224
QY -234 I--DIVDPFLTLKGL 246
Db 225 IKHEIFDEDLTIKGV 239

RESULT 12 .
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C;Accession: D71326
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDaniel, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: D71326
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <COL>
A;Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g3322717
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0431
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 24.5%; Score 327.5; DB 2; Length 273;
Best Local Similarity 29.9%; Pred. No. 1.3e-20;
Matches 75; Conservative 64; Mismatches 105; Indels 7; Gaps 4;
QY 2 LLLVIDVGNNTVLGVY--HDGK--LEYHWRHRIETSRHKTEDEFGMILRSLFDHSGLMFEQI 57
Db 1 MLLIDVGNSHVVFQIGENGGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRASL 60

QY 58 DGIIISSVPPIMPFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNP--KEVCADRIVNAV 115
Db 61 RDAFISSVVPVLTKTIADAVAQISGVQVPVFGPWAYEHLVPRIPEPVRAEIGTDLVANAV 120
QY 116 AAHLVGNPLIVVDFGTATTTCYIDENKQVMGGAIAPGITISTEALYSRAAKLPRIEITR 175
Db 121 AAYHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTGAQLPLVPLAL 180
QY 176 PDNIIGKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCID 235
Db 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRIFSSEVD-FP 239
QY 236 IVDPPFLTLKGL 246
Db 240 PIDAQLTLTSLGL 250

RESULT 13
F70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: F70165
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: F70165
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-262 <KLE>
A;Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g2688431
A;Experimental source: strain B31
Query Match 20.3%; Score 271.5; DB 2; Length 262;
Best Local Similarity 27.3%; Pred. No. 7.9e-16;
Matches 70; Conservative 57; Mismatches 108; Indels 21; Gaps 5;
QY 3 LVLDVGNNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62
Db 9 LIIDIGNTSIAFALKDNQVNLFIKMKTNLMRLYDEVYSFEEFNFDN-----VNKVFI 62
QY 63 SSVVPPIMPFALERMCTKYFHIEPQIVGPGMKTGLNIKYD---NPKE-----VGADRIVNA 114
Db 63 SSVVPIILNETFKNVIFSFFKIKPLFI-----GFDLNYDLTFNYPYKSKFLGSDVFANL 116
QY 115 VAAIHLVGNPLIVVDFGTATTTCYIDENKQVMGGAIAPGITISTEALYSRAAKLPRIEI 173
Db 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGLGGIINSGLINFSNLLDNYAYLIKFPPI 176
QY 174 TRPDNIIGKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDC 233
Db 177 STPNLLERTTSGSVNSGLFYQYKYLIEGVYVDIKQMYKKFNLIITGGNADLILSLIEI 236
QY 234 IDIVDPFLTLKGLELI 249
Db 237 EFIFNIHLTVEGVRIL 252

RESULT 14
AI2292
hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AI2292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana-

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:59:29 ; Search time 40 Seconds
(without alignments)
1664.439 Million cell updates/sec

Title: US-09-813-453B-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1036	77.6	258	2 Q9F985	Q9f985 bacillus st
2	987	73.9	254	16 Q9KGH5	Q9kgh5 bacillus ha
3	853.5	63.9	255	16 Q8EU15	Q8eul5 oceanobacil
4	852	63.8	259	16 Q8YAC5	Q8yac5 listeria mo
5	847	63.4	259	16 Q92F54	Q92f54 listeria in
6	846	63.4	259	2 Q8KU01	Q8ku01 listeria mo
7	775	58.1	273	16 Q97EB4	Q97eb4 clostridium
8	773	57.9	255	16 Q8R7M2	Q8r7m2 thermoanaer
9	737	55.2	259	16 Q8XHL5	Q8xhl5 clostridium
10	655	49.1	265	16 Q9X8N6	Q9x8n6 streptomyc
11	577	43.2	261	16 Q9A6Z1	Q9a6z1 caulobacter
12	539.5	40.4	256	16 Q8G558	Q8g558 bifidobacte
13	519.5	38.9	274	16 Q9CD56	Q9cd56 mycobacteri
14	503.5	37.7	257	16 Q8F7V9	Q8f7v9 leptospira
15	494.5	37.0	272	16 O06282	O06282 mycobacteri
16	458	34.3	256	16 Q8RFE4	Q8rfe4 fusobacteri

17	428	32.1	262	16 Q9RX54	Q9rx54 deinococcus
18	405.5	30.4	246	16 Q9WZY5	Q9wzy5 thermotoga
19	338.5	25.4	212	2 Q32514	Q32514 desulfovibr
20	327.5	24.5	273	16 O83446	O83446 treponema p
21	319	23.9	270	16 O8KCK7	O8kck7 chlorobium
22	271.5	20.3	262	16 O51477	O51477 borrelia bu
23	185.5	13.9	255	16 Q8EUB0	Q8eub0 mycoplasma
24	171	12.8	295	16 Q8Y2M4	Q8y2m4 ralstonia s
25	170	12.7	276	16 Q8YQD7	Q8yqgd7 anabaena sp
26	166	12.4	56	2 P94305	P94305 bacillus ps
27	166	12.4	254	16 Q8DJS3	Q8djs3 synechococc
28	163	12.2	257	16 P74045	P74045 synechocyst
29	154.5	11.6	229	16 O67753	O67753 aquifex ao
30	150	11.2	592	16 Q9JXF1	Q9jxf1 neisseria m
31	150	11.2	592	16 Q9JWI7	Q9jwi7 neisseria m
32	134.5	10.1	242	16 Q9PCI4	Q9pci4 xylella fas
33	133	10.0	248	16 Q9HWC1	Q9hwc1 pseudomonas
34	119.5	9.0	242	16 Q8PFG5	Q8pfg5 xanthomonas
35	117	8.8	224	16 Q98Q93	Q98q93 mycoplasma
36	116.5	8.7	223	16 Q9ZKY6	Q9zky6 helicobacte
37	116.5	8.7	242	16 Q8P3Y1	Q8p3y1 xanthomonas
38	109.5	8.2	223	16 O25533	O25533 helicobacte
39	109	8.2	209	16 Q9PIA9	Q9pia9 campylobact
40	104.5	7.8	597	17 O30225	O30225 archaeoglob
41	100	7.5	318	10 Q9LVY3	Q9lv3 arabidopsis
42	95	7.1	520	16 Q8U6P6	Q8u6p6 agrobacteri
43	94.5	7.1	858	12 Q9QTD8	Q9qtd8 marek's dis
44	93.5	7.0	895	17 Q972N0	Q972n0 sulfolobus
45	93	7.0	676	12 Q8JTV2	Q8jtv2 lumpy skin

ALIGNMENTS

RESULT 1

Q9F985 ID Q9F985 PRELIMINARY; PRT; 258 AA.
AC Q9F985;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Putative 32 kDa replication protein.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V;
RA Vasquez C., Pichuanes S., Saavedra C.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF198621; AAG28531.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 77.6%; Score 1036; DB 2; Length 258;
Best Local Similarity 77.9%; Pred. No. 1.3e-85;
Matches 197; Conservative 31; Mismatches 25; Indels 0; Gaps 0;

QY	1	LLLVIDVGNNTVLGVYHDGKLEYHWRITSRHKTDEFGMILRSLFDHSGLMFEQIDGI 60	
Db	1	MIFVLDVGNNTVLGVYDGDDELKHHWRITSRGKTEDEYGMTIKALLNHVGLQFSDIDGI 60	
QY	61	IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNLIKNDPNKVEGADRIVNAVAIHL 120	
Db	61	IISSVVPPIMFALERMCLKYFHIKPIIVGPGIKTGLNLIKNDPNPREVGADRIVNAVAGIHL 120	
QY	121	YGNPLIVVDFGTATTTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIETRPDNI 180	
Db	121	YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPGIMISTEALFARAAKLPRIETARPDDI 180	
QY	181	GKNTVSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGTGLAPLIANESDCIDIVDPF 240	

Db 181 GKNVTSAMQAGILGYGVQVEGIVSRMKAKSPVPPKVIATGGLASLIASESNVIDIVDPF 240
241 LTLKGLLELIYERN 253
241 LTLTGLKILYEKN 253

RESULT 2

Q9KGHS PRELIMINARY; PRT; 254 AA.
AC Q9KGHS;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.,
RA "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL:# AP001507; BAB03805.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 73.9%; Score 987; DB 16; Length 254;
Best Local Similarity 72.3%; Pred. No. 3.4e-81;
Matches 183; Conservative 33; Mismatches 37; Indels 0; Gaps 0;

QY 1 LLLVIDGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60
Db 1 MILVIDGNTNTVLGVYQDETTLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
QY 61 IISSVVPPIPFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120
Db 61 VISSVPPMFMFSLQCMCKKYFHVTPMIGIGIKTGLNIKYDNPKEVGADRIVNAVAIEL 120
QY 121 YGNPLIVDFGTATTYCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGPAIVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQVV 180
QY 181 GKNTVSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GTNTIDSMQSGIFYGVYVQVDGVVVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSF 240
QY 241 LTLKGLLELIYERN 253
Db 241 LTLKGLQLIYKKN 253

RESULT 3

Q8EU15 PRELIMINARY; PRT; 255 AA.
AC Q8EU15;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical conserved protein.
GN OB0081.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.

OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004593; BAC12037.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 255 AA; 28159 MW; 01A338CFF0F70267 CRC64;
Query Match 63.9%; Score 853.5; DB 16; Length 255;
Best Local Similarity 63.0%; Pred. No. 4.2e-69;
Matches 160; Conservative 42; Mismatches 51; Indels 1; Gaps 1;

QY 1 LLLVIDGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60
Db 1 MLFVLDVGNTNTVLGVFDQGELTHHWRIKTDYKTEDEFGMLIHSFLQHKLSFEDIKGV 60
QY 61 IISSVVPPIPFALERMCTKYFHIEPQIVG-PGMKTGLNIKYDNPKEVGADRIVNAVAIHL 119
Db 61 IISSVVPPIPFALERMKSRDYFHLDAVIGKTSYQTFLNMYPNPQEIAGADRIVNAVAATE 120
QY 120 LYGNPLIVDFGTATTYCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETRPDNI 179
Db 121 EYGSPLIIIDFGTATTYCYIDEDTAYAGGIITPGINISMEALYSNASKLPKIEIQKETV 180
QY 180 IGKNTVSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDP 239
Db 181 IGSTTVDMTSGVFGYIGQVDGLVKRIKDEKGTNFTVIATGGLAKLIAHESASIDIVEP 240
QY 240 FLTLKGLLELIYERN 253
Db 241 YLTLKGLHLIYQKN 254

RESULT 4

Q8YAC5 PRELIMINARY; PRT; 259 AA.
ID Q8YAC5
AC Q8YAC5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madsen E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591974; CAD00748.1; -.
DR MEROPS; M41.009; -.
DR ListiList; LMO00221; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.

DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 63.8%; Score 852; DB 16; Length 259;
Best Local Similarity 63.4%; Pred. No. 5.8e-69;
Matches 161; Conservative 37; Mismatches 56; Indels 0; Gaps 0;

QY 1 LLLVIDVGNNTVLGVYHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MILVIDVGNNTCTGVYKEQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

QY 61 IISSVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120
Db 61 IISSVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

QY 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGTPVIVDFGTATTCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPVVDIAESSQII 180

QY 181 GKNVTSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKSTVSSMQAGIFYGFGQCEGIIAEMKKQSNASPVVATGGLARMITEKSSAVDILDPP 240

QY 241 LTLKGLELIYERNR 254
Db 241 LTLKGLELLYRRNK 254

RESULT 5
Q92F54 ID Q92F54 PRELIMINARY; PRT; 259 AA.
AC Q92F54;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lin0253.
GN LIN0253.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11679669;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A., Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T., Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P., Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O., Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P., Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G., Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H., Nordisiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R., Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A., Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.; "Comparative genomics of Listeria species."; Science 294:849-852(2001).
RL EMBL; AL596164; CAC95486.1; --
DR ListiList; LIN00253; --
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 63.4%; Score 847; DB 16; Length 259;
Best Local Similarity 62.6%; Pred. No. 1.6e-68;
Matches 159; Conservative 39; Mismatches 56; Indels 0; Gaps 0;

QY 1 LLLVIDVGNNTVLGVYHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MILVIDVGNNTCTGVYKEQKLLRHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

QY 61 IISSVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120
Db 61 IISSVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

QY 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGTPVIVDFGTATTCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPVVDIAESSQII 180

QY 181 GKNVTSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKSTVASMQAGIFYGFGQCEGIIAEMKKQSNTPSPVVATGGLARMITEKSSAVDILDPP 240

QY 241 LTLKGLELIYERNR 254
Db 241 LTLKGLELLYRRNK 254

RESULT 6
Q8KU01 ID Q8KU01 PRELIMINARY; PRT; 259 AA.
AC Q8KU01;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE YacB-like protein.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4b1;
RA Li G., Kathariou S.; Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF467001; AAM74003.1; --
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
SQ SEQUENCE 259 AA; 28180 MW; 6810695C23F25DEC CRC64;

Query Match 63.4%; Score 846; DB 2; Length 259;
Best Local Similarity 62.6%; Pred. No. 2e-68;
Matches 159; Conservative 40; Mismatches 55; Indels 0; Gaps 0;

QY 1 LLLVIDVGNNTVLGVYHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
Db 1 MILVIDVGNNTCTGVYKEQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

QY 61 IISSVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120
Db 61 IISSVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

QY 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIETRPDNI 180
Db 121 YGTPVIVDFGTATTCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPVVDIAESNQII 180

QY 181 GKNVTSAMQSGILFGYVQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
Db 181 GKSTVASMQAGIFYGFGQCEGIIAEMKKQSNASPVVATGGLARMITEKSSAVDILDPP 240

QY 241 LTLKGLELIYERNR 254
Db 241 LTLKGLELLYRRNK 254

RESULT 7
Q97EB4 ID Q97EB4 PRELIMINARY; PRT; 273 AA.
AC Q97EB4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory

Db 61 IISVVPPNIMHSLENNVRKCFCKEPIVVGPGIKTGINIKYDNPKEVGADRIVNAVAAFEK 120

QY 121 YGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETRPDNI 180

Db 121 HKKPMIIDDFTATTTCALTEKGDYLGNCICPGIQISADALFERAAKLPRIELEKPKSVI 180

QY 181 GKNTVSAMQSGILFGYVGVQVEGIVKRMKWQ----AKQDLKVIATGGLAPLIANESDCIDI 236

Db 181 CKNTVTSMQAGIIYGYIGKVEYIVKRMKEMMDLGEKEPFFVLATGGLAKLVYSETDVIDE 240

QY 237 VDPFLLTKGLELIYERNR 254

Db 241 VDRKLTLEGLKILYKKNK 258

RESULT 10

Q9X8N6 PRELIMINARY; PRT; 265 AA.

AC Q9X8N6; 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Hypothetical protein SCO3380.

GN SCO3380 OR SCE94.31C.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1902;

RN [1] SEQUENCE FROM N.A.

RP STRAIN=A3(2);

RC Oliver K., Harris D.; Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN [2] SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.; Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN [3] SEQUENCE FROM N.A.

RP STRAIN=A3(2);

RC MEDLINE=97000351; PubMed=8843436;

RX Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;

RA "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

RN [4] SEQUENCE FROM N.A.

RP STRAIN=A3(2) / M145;

RC MEDLINE=21996410; PubMed=12000953;

RX Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S., Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A.;

RA "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).";

RL Nature 417:141-147(2002).

DR EMBL; AL939116; CAB40880.1; -

DR InterPro; IPR004619; Baf.

DR Pfam; PF03309; Bvg_acc_factor; 1.

DR TIGRFAMs; TIGR00671; baf; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 49.1%; Score 655; DB 16; Length 265;

Best Local Similarity 51.0%; Pred. No. 4.1e-51;

Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3;

QY 1 LLLVIDVGNNTVVGVDGKLEYHWRITSRHKTEDEFGMILRSFDHSGLMFEQ---- 56

Db 1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRRTADELAVLLQGLMGHPLGLGDELGDG 60

QY 57 IDGIISSVPPINFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115

Db 61 IDGIAICATVPSVLHELREVTTRYGVDPAVLVEPGVKTGPILTDPKEVGADRIINAV 120

QY 116 AAHLGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETR 175

Db 121 AAVELYGGPAIVVDFGTATTFDVARSARGEYIGGVIAPIGIEISVEALGVKGALRKIEVAR 180

QY 176 PDNIIGKNTVSAMQSGILFGYVGVQVEGIVKRMKWQ---KQDLKVIATGGLAPLIANESD 232

Db 181 PRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDDPDDVTVTIATGGLAPMWLGESS 240

QY 233 CIDIVDPFLLTKGLELIYERN 253

Db 241 VIDEHEPWLTLMGLRLVYERN 261

RESULT 11

Q9A6Z1 PRELIMINARY; PRT; 261 AA.

AC Q9A6Z1; 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Transcriptional activator, putative, Baf family.

GN CC1935.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;

OC Caulobacteraceae; Caulobacter.

OX NCBI_TaxID=155892;

RN [1] SEQUENCE FROM N.A.

RP STRAIN=ATCC 19089 / CB15;

RC MEDLINE=21173698; PubMed=11259647;

RX Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL; AE005867; AAK23910.1; -

DR TIGR; CC1935; -

DR InterPro; IPR004619; Baf.

DR Pfam; PF03309; Bvg_acc_factor; 1.

DR TIGRFAMs; TIGR00671; baf; 1.

KW Complete proteome.

SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 43.2%; Score 577; DB 16; Length 261;

Best Local Similarity 44.4%; Pred. No. 4.6e-44;

Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;

QY 1 LLLVIDVGNNTVVGVDGKLEYHWRITSRHKTEDEFGMILRSFDHSGLMFEQIDGI 60

Db 2 MLLAIEQGNNTMTMFAIHGASWVAQWRSATESSTRTADEVVWLSQLLSMQGLGFRDAIDAV 61

QY 61 IISVVPPINFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120

Db 62 IISVVVPQISIFNLRLNLSRRYFNVEPLVIGENAKLIGIDVRIEKPSEAGADRLVNAIGAAMV 121

QY 121 YGNPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETRP--DN 178

Db 122 YPGPLVIDSGTATTFDIVAADGAFFEGGIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 181

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2003, 12:44:12 ; Search time 21 Seconds
(without alignments)
519.819 Million cell updates/sec

Title: US-09-813-453B-2
Perfect score: 1335
Sequence: 1 LLLVDVGNNTNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181.5	13.6	249	4	US-09-328-352-4748
2	133	10.0	253	4	US-09-252-991A-21337
3	88	6.6	311	4	US-09-252-991A-23527
4	86	6.4	1011	4	US-09-252-991A-22505
5	83.5	6.3	207	3	US-08-549-515-9
6	80	6.0	449	4	US-09-530-836-2
7	80	6.0	502	4	US-09-252-991A-27619
8	79	5.9	252	4	US-09-328-352-5774
9	79	5.9	342	3	US-08-096-181A-12
10	79	5.9	342	3	US-08-096-181A-14
11	79	5.9	342	5	PCT-US94-08326-12
12	79	5.9	342	5	PCT-US94-08326-14
13	79	5.9	361	3	US-08-096-181A-8
14	79	5.9	361	5	PCT-US94-08326-8
15	79	5.9	363	3	US-08-096-181A-10
16	79	5.9	363	5	PCT-US94-08326-10
17	79	5.9	923	3	US-08-936-135-6
18	78.5	5.9	357	4	US-09-120-051D-9
19	78	5.8	1012	1	US-07-944-943-2
20	78	5.8	1012	1	US-08-219-262B-2
21	78	5.8	1012	3	US-09-031-655-2
22	78	5.8	1213	2	US-08-937-102-2
23	78	5.8	3898	2	US-08-876-991-2
24	78	5.8	3898	2	US-09-059-853-2
25	78	5.8	3898	3	US-08-750-717-2
26	77.5	5.8	326	4	US-09-328-352-5506
27	77.5	5.8	1107	4	US-09-358-383C-16

28	77	5.8	627	4	US-09-071-035-452	Sequence 452, App
29	77	5.8	1313	4	US-09-071-035-450	Sequence 450, App
30	77	5.8	1313	4	US-09-071-035-454	Sequence 454, App
31	76	5.7	922	4	US-09-116-473-4	Sequence 4, Appli
32	75.5	5.7	443	4	US-09-594-193-9	Sequence 9, Appli
33	75.5	5.7	551	4	US-09-328-352-7176	Sequence 7176, Ap
34	75.5	5.7	1215	4	US-09-134-001C-5319	Sequence 5319, Ap
35	75	5.6	553	6	5310678-1	Patent No. 5310678
36	75	5.6	1012	1	US-08-219-262B-9	Sequence 9, Appli
37	75	5.6	1012	3	US-09-031-655-9	Sequence 9, Appli
38	74.5	5.6	246	4	US-09-634-238-224	Sequence 224, App
39	74.5	5.6	262	4	US-09-252-991A-21196	Sequence 21196, A
40	74.5	5.6	452	4	US-09-252-991A-31196	Sequence 31196, A
41	74.5	5.6	457	4	US-09-134-001C-3838	Sequence 3838, Ap
42	74	5.5	413	3	US-09-147-928-2	Sequence 2, Appli
43	73.5	5.5	416	4	US-09-328-352-6675	Sequence 6675, Ap
44	73.5	5.5	473	4	US-09-107-532A-5371	Sequence 5371, Ap
45	73.5	5.5	499	4	US-09-328-352-5089	Sequence 5089, Ap

ALIGNMENTS

RESULT 1
US-09-328-352-4748
; Sequence 4748, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4748
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4748

Query Match	13.6%	Score	181.5	DB	4	Length	249
Best Local Similarity	27.6%	Pred. No.	4.8e-13				
Matches	79	Conservative	39	Mismatches	73	Indels	95
Gaps	18						
QY	3	LVIDVGNNTNTVLGVYHDGKLEYHWRIETSR	-----HKTEDE-----	FGMILRSFLPDHSG	51		
Db	9	LWLDIGNT-----RLKY-WITENOQIIEHAAELHLOSPADLLGLIQH--FKHQG	55				
QY	52	LMFEQIDGIIISVVVP-----PIMPALERMCTKYFHIEPQIVGPGMKTG	95				
Db	56	L-----HRIGISSVLDTENNQRIQILKWLEIPVVFVFA--KVHAEY-----	AG	95			
QY	96	LNKYDNPKVGDRIVNAVAIAHLYGNPLIVVDFGTATTYCYID--ENKQYMGGAIA	PG	153			
Db	96	LQCGYEVPSQLGIDRWLQVLAAVEEKEN-YCIIGCGTALT---IDLTKGKOHLLGGVILR	151				
QY	154	ITISTEALY--SRAAKLPRIETRPDNI-IGKNTVSAMQSGILFGYVGQVEGIVKRMKWQ	210				
Db	152	LYLQORDALIQNTGKIKIPDSAF---DNLNPENNTVDVHHGILLGLISTIESIMQ-----	203				
QY	211	AKQDLKVIATGGLAPLIAN-----ESDCIDIVDPFLTLKGLE	247				
Db	204	-QSPKLLLTGGDAPLFAKFLQKYQPTVETD-----LLLKGLQ	240				

RESULT 2
US-09-252-991A-21337
; Sequence 21337, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-549-515-9

Query Match 6.3%; Score 83.5; DB 3; Length 207;
Best Local Similarity 24.4%; Pred. No. 0.14;
Matches 47; Conservative 27; Mismatches 66; Indels 53; Gaps 11;

QY 15 GYVHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQIDGIIISSVVPIMFALE 74
Db 52 GWHQNVFAYY--LSISCNHCEDP---ACTKVCPSGAMHKREDGFVVVD-----E 96

QY 75 RMC--TKYFHIEPQIVGPGMKTGLNIKYDNPK-----EVGADRIVNAVAIAHLYGNPLI 126
Db 97 DVCIGCRYCHMACPYGAP-----QYNETKGHMTKDCGCDYDRAEGKKPICVESCPLR 148

QY 127 VDFGTATTYCYIDENKQYMG--GAIAPGITISTEALYSRAAKLPRIETRPDNIIGKNT 184
Db 149 ALDFGP-----IDELRKKHGDLAAVAP-----RA--LPRAHFTKP-NIVIKPN 188

QY 185 VSAMQSGILFGYV 197
Db 189 ANSRPTGDTTGYL 201

RESULT 6
US-09-530-836-2
; Sequence 2, Application US/09530836
; Patent No. 6534284
; GENERAL INFORMATION:
; APPLICANT: Mohammed El-Sherbeini
; APPLICANT: Kenny Kin Wong
; APPLICANT: Wayne M. Geissler
; TITLE OF INVENTION: MURD PROTEIN AND GENE OF STAPHYLOCOCCUS
; FILE REFERENCE: 19985P
; CURRENT APPLICATION NUMBER: US/09/530,836
; CURRENT FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 60/064,183
; PRIOR FILING DATE: 1997-11-04
; PRIOR APPLICATION NUMBER: PCT/US98/23156
; PRIOR FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Staphylococcus aureus Murd
US-09-530-836-2

Query Match 6.0%; Score 80; DB 4; Length 449;
Best Local Similarity 26.9%; Pred. No. 1.2;
Matches 35; Conservative 21; Mismatches 42; Indels 32; Gaps 6;

QY 93 KTGLNIKYDNPKEVGADRI-----VNAVAIAH-----LYGNPLIVVDFGTATTYCYID 140
Db 30 KLGANVTVNDGKDLSDAHAKDLESMGISVSGSHPLTLTDNNPIIVKNPGIPYTVSIID 89

QY 141 ENKQYMGGAIPGITISTEA-LYSRAAKLPRIETRPDNIIGKNTVSAM-----QS 190

Db 90 E-----AVKRGKILTEVELSYLISEAPIIATVTGN---GKTTVTSLIGDMFKKSRLT 139

QY 191 GILFGYVGQV 200
Db 140 GRLSGNIGYV 149

RESULT 7
US-09-252-991A-27619
; Sequence 27619, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27619
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27619

Query Match 6.0%; Score 80; DB 4; Length 502;
Best Local Similarity 23.2%; Pred. No. 1.5;
Matches 33; Conservative 20; Mismatches 53; Indels 36; Gaps 6;

QY 19 DGKLEYHWRITSRHKTEDEFGMILRSLFDHSGLM-----FEQIDGIISSWVPP 68
Db 258 EGMQHRHRDRVEGR--AEDHSGEVLRRLAQHRLDHEAQGAFADEQLAHIVAGGVLDQ 315

QY 69 IMFALERMCTKYFHIEP--QIVGPGMKTGLNIKYDNPKEVGAD-----RIV 112
Db 316 VLVQLEQVALAGDDLHPGHPVTGHAVANDL-----DPAGIGADIAADLAGTGRSEVNRV 370

QY 113 NAV--AAIHLNGNPLIVVDFG 131
Db 371 QAVFLGEGQLLGNHARLADYG 392

RESULT 8
US-09-328-352-5774
; Sequence 5774, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5774
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5774

Query Match 5.9%; Score 79; DB 4; Length 252;
Best Local Similarity 21.1%; Pred. No. 0.65;
Matches 55; Conservative 49; Mismatches 83; Indels 74; Gaps 16;

QY 7 VGNTNTVLGVYHDGKLEY---HWR-----IETSRHKTEDE-----F 39
Db 17 IDSIRTVL--FFNGEINFKKKEWSKNITGLEISNEMTQSEENRLIQYVEITNLDNKNQF 74

QY 40 GMIL---RSLFDHSGLMFEQIDGI-IISSVVPPIPFALERMCTKYFHIEPQIVGPGMKTG 95

Db 75 NLVYLKQSLIDLQ-LVFERDENFYTFENEIIKEVDFFYEXISVFFDQLNEKIIRIGNVVE 133

Qy 96 LNIKYDNPKEVGADRIVNAVAIAHLYGNPLIVVDFGTATTCYIDENK-----QYMGG-- 148

Db 134 LSIKPVNEK-IGCDLLRSNVSYLNNMQEDLEEISYRTNKSQ-FIDNIKINQVVQYSGNQK 191

Qy 149 ---AIAFGITISTEALYSRAAKLPRIEITRPDNI---IGKNTVSAMQSGILF-GYVGQVE 201

Db 192 MSLVIDPNIGI-----PKAKVQK--NILMNIDVNTDASHRSELDLFLKFIPLLQ 237

Qy 202 GIVKRMKQAKQDLKVIATGG 222

Db 238 DSVK-----KLIRNGG 248

RESULT 9

US-08-096-181A-12

; Sequence 12, Application US/08096181A

; Patent No. 6153406

; GENERAL INFORMATION:

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Pullen, Jeffrey K.

; APPLICANT: Soper, Thomas S.

; APPLICANT: Liang, Shu-Mei

; TITLE OF INVENTION: A Method For The High Level Expression, Purification And Refolding Of The Outer Membrane Protein

; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein

; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/096,181A

; FILING DATE: 23-Jul-1993

; CLASSIFICATION: 424

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 342 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-096-181A-12

Query Match 5.9%; Score 79; DB 3; Length 342;

Best Local Similarity 20.1%; Pred. No. 1.1;

Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

Qy 28 IETSRHKT-----EDEFGMILRSLF-----DHSGLMFEQIDGIISSVVPIMFAL 73

Db 100 VKLGRAKTIADGITS AEDKEYGVLNNSDIPTSGNTVGYTFKGDGLVGA---NYLLAQ 156

Qy 74 ERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIAHLYGNPLIVVDFGTA 133

Db 157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 203

Qy 134 TTYCY--IDENKQYMGGAIA-----GITISTEALYSRAAKLP-----R 170

Db 204 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 263

Qy 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195

RESULT 11

PCT-US94-08326-12

; Sequence 12, Application PC/TUS9408326

; GENERAL INFORMATION:

; APPLICANT: North American Vaccine, Inc.

; APPLICANT: 12103 Indian Creek Court

; APPLICANT: Beltsville, MD 20705

; APPLICANT: Pullen, Jeffrey K.

Db 264 YELMEDTNVYGNFKYERTSVDOGEKTRQAVLFG 297

RESULT 10

US-08-096-181A-14

; Sequence 14, Application US/08096181A

; Patent No. 6153406

; GENERAL INFORMATION:

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Pullen, Jeffrey K.

; APPLICANT: Soper, Thomas S.

; APPLICANT: Liang, Shu-Mei

; TITLE OF INVENTION: A Method For The High Level Expression, Purification And Refolding Of The Outer Membrane Protein

; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein

; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/096,181A

; FILING DATE: 23-Jul-1993

; CLASSIFICATION: 424

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 342 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-096-181A-14

Query Match 5.9%; Score 79; DB 3; Length 342;

Best Local Similarity 20.1%; Pred. No. 1.1;

Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

Qy 28 IETSRHKT-----EDEFGMILRSLF-----DHSGLMFEQIDGIISSVVPIMFAL 73

Db 100 VKLGRAKTIADGITS AEDKEYGVLNNSDIPTSGNTVGYTFKGDGLVGA---NYLLAQ 156

Qy 74 ERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIAHLYGNPLIVVDFGTA 133

Db 157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 203

Qy 134 TTYCY--IDENKQYMGGAIA-----GITISTEALYSRAAKLP-----R 170

Db 204 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 263

Qy 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195

Db 264 YELMEDTNVYGNFKYERTSVDOGEKTRQAVLFG 297

RESULT 11

PCT-US94-08326-12

; Sequence 12, Application PC/TUS9408326

; GENERAL INFORMATION:

; APPLICANT: North American Vaccine, Inc.

; APPLICANT: 12103 Indian Creek Court

; APPLICANT: Beltsville, MD 20705

; APPLICANT: Pullen, Jeffrey K.

APPLICANT: Soper, Thomas S.
 APPLICANT: Liang, Shu-Mei
 TITLE OF INVENTION: A Method For The High Level
 TITLE OF INVENTION: Expression,
 TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
 TITLE OF INVENTION: Protein
 TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 STREET: 1100 New York Avenue, Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/08326
 FILING DATE: Herewith
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/096,181
 FILING DATE: 23-JULY-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REFERENCE/DOCKET NUMBER: 1438.001PC01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 342 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-08326-12

```

; TITLE OF INVENTION: Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
; TITLE OF INVENTION: Protein
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08326
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,181
; FILING DATE: 23-JULY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REFERENCE/DOCKET NUMBER: 1438.001PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-08326-14

Query Match 5.9%; Score 79; DB 5; Length 342;
Best Local Similarity 20.1%; Pred. No. 1.1;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps

QY 28 IETSRHKT-----EDEFGMILRSLF-----DHSGLMFEQIDGIISSVWPPIMFAL
Db 100 VKIGRAKTIADGITS AEDKEYGV LNNSDYIPTSGNTVGYTFKIGDGLVGA---NYLLAQ
QY 74 ERMCTKYFHIEPQIVGPGMKTGLN KYDNPKEVGADRI VNAVAAIHLYGNPLIVVDFGTA
Db 157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR-----
QY 134 TTYCY--IDENKQYMGCAIAP-----GITISTEALYSRAAKLP-----R
Db 204 TNYKYNESDEHKQQLNGCVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFG
QY 171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195
Db 264 YELMEDTNYVGNFKYERTSV DQGEKTREQAVLFG 297

RESULT 13
US-08-096-181A-8
; Sequence 8, Application US/08096181A
; Patent No. 6153406
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

```

```

RESULT 12
PCT-US94-08326-14
; Sequence 14, Application PC/TUS9408326
; GENERAL INFORMATION:
; APPLICANT: North American Vaccine, Inc.
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, MD 20705
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level

```


ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,181A
FILING DATE: 23-Jul-1993
CLASSIFICATION: 424
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-181A-8

Query Match 5.9%; Score 79; DB 3; Length 361;
Best Local Similarity 20.1%; Pred. No. 1.2;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

Qy 28 IETSRHKT-----EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIPFAL 73
Db 119 VKLGRAKTIADGITS AEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 175
Qy 74 ERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIAHLYGNPLIVVDFGTA 133
Db 176 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 222
Qy 134 TTYCY--IDENKQYMGGAIAIP-----GITISTEALYSRAAKLP-----R 170
Db 223 TNYKYNESDEHKQQLNGVLATLGYRPSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 282
Qy 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195
Db 283 YELMEDTNVYGNFKYERTSVDDQGEKTRQAVLFG 316

RESULT 14
PCT-US94-08326-8
Sequence 8, Application PC/TUS9408326
GENERAL INFORMATION:
APPLICANT: North American Vaccine, Inc.
APPLICANT: 12103 Indian Creek Court
APPLICANT: Beltsville, MD 20705
APPLICANT: Pullen, Jeffrey K.
APPLICANT: Soper, Thomas S.
APPLICANT: Liang, Shu-Mei
TITLE OF INVENTION: A Method For The High Level
TITLE OF INVENTION: Expression,
TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
TITLE OF INVENTION: Protein
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/08326
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,181
FILING DATE: 23-JULY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REFERENCE/DOCKET NUMBER: 1438.001PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-08326-8

Query Match 5.9%; Score 79; DB 5; Length 361;
Best Local Similarity 20.1%; Pred. No. 1.2;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

Qy 28 IETSRHKT-----EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIPFAL 73
Db 119 VKLGRAKTIADGITS AEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 175
Qy 74 ERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIAHLYGNPLIVVDFGTA 133
Db 176 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR----- 222
Qy 134 TTYCY--IDENKQYMGGAIAIP-----GITISTEALYSRAAKLP-----R 170
Db 223 TNYKYNESDEHKQQLNGVLATLGYRPSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 282
Qy 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195
Db 283 YELMEDTNVYGNFKYERTSVDDQGEKTRQAVLFG 316

RESULT 15
US-08-096-181A-10
Sequence 10, Application US/08096181A
Patent No. 6153406
GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
APPLICANT: Pullen, Jeffrey K.
APPLICANT: Soper, Thomas S.
APPLICANT: Liang, Shu-Mei
TITLE OF INVENTION: A Method For The High Level Expression,
TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,181A
FILING DATE: 23-Jul-1993
CLASSIFICATION: 424

; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

LENGTH: 363 amino acids

TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-096-181A-10

Query Match	5.9%;	Score 79;	DB 3;	Length 363;
Best Local Similarity	20.1%;	Pred. No. 1.2;		
Matches 43;	Conservative 38;	Mismatches 71;	Indels 62;	Gaps 10;

Qy	28	IETSRHKT-----EDEFGMILRSLF-----DHSGLMFEQIDGIIISVVPPIMFAL 73
Dd	121	VKLGRAKTIADGITSAEDEKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 177
Qy	74	ERMCTKYFHIEPIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVDFGTA 133
Dd	178	KREGAKGENKRPNDKACEVRIG--EINNGIQVGAKYDANDIVAKIAYGR----- 224
Qy	134	TTYCY--IDENKQYMCGAIAP-----GITISTEALYSRAAKLP-----R 170
Dd	225	TNYKYNESDEHKQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 284
Qy	171	IEITRPDNIIG-----KNTV----SAMQSGLIFG 195
Dd	285	YELMEDTNNVGNFKYERTSDVQQEKTRQEAVLFG 318

Search completed: December 18, 2003, 12:56:17
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: December 18, 2003, 12:56:23 ; Search time 17 Seconds
(without alignments)
713.699 Million cell updates/sec
Title: US-09-813-453B-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNTNTVLGVYHDG.....PFLTCLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1114	83.4	233	1	YACB_BACSU	P37564 bacillus su
2	129.5	9.7	267	1	BAF_BORPE	Q45338 bordetella
3	90.5	6.8	4451	1	GRSB_BACBR	P14688 b grammicidi
4	89.5	6.7	357	1	BUK_THETN	Q8r832 thermoanaer
5	88.5	6.6	425	1	PUR_A_LEPIN	Q8f738 leptospira
6	88	6.6	256	1	HI61_PSEAE	Q9hu44 pseudomonas
7	87.5	6.6	424	1	SAHH_METKA	P58855 methanopyru
8	87	6.5	336	1	PYRD_SALTI	Q82789 salmonella
9	87	6.5	336	1	PYRD_SALTY	P25468 salmonella
10	86.5	6.5	243	1	Y004_METJA	Q60315 methanococc
11	85	6.4	336	1	PYRD_ECOL6	Q8fj91 escherichia
12	85	6.4	336	1	PYRD_ECOLI	P05021 escherichia
13	83.5	6.3	204	1	DMSB_ECOLI	P18776 escherichia
14	83	6.2	187	1	ATKC_RHILO	Q98gx7 rhizobium 1
15	83	6.2	208	1	PIMT_VIBPA	Q871q6 vibrio para
16	83	6.2	300	1	NANK_HAEIN	P44541 haemophilus
17	83	6.2	425	1	GSA_CLOPE	Q9znc8 clostridium
18	83	6.2	470	1	CH60_TRIVA	Q95058 trichomonas
19	83	6.2	628	1	GATE_SULTO	Q971w7 sulfolobus
20	82	6.1	305	1	GALE_METJA	Q57664 methanococc
21	82	6.1	3898	1	POLG_HCVB	P21530 hog cholera
22	81	6.1	359	1	OP21_HAEIN	P43839 haemophilus
23	81	6.1	449	1	MURD_STAAN	Q8nx35 staphylococ
24	81	6.1	1224	1	RPB2_YEAST	P08518 saccharomyc
25	80.5	6.0	208	1	PIMT_YERPE	Q8zbg0 yersinia pe
26	80.5	6.0	297	1	NANK_PASMU	Q9ckb3 pasteurella
27	80.5	6.0	327	1	EBGR_ECOLI	P06846 escherichia
28	80.5	6.0	354	1	LUXA_VIBFI	P19907 vibrio fisc
29	80.5	6.0	410	1	RAA2_CHLRE	Q9smh4 chlamydomon
30	80	6.0	199	1	ATC2_ANASP	Q8ypf1 anabaena sp
31	80	6.0	449	1	MURD_STAAM	O33595 staphylococ
32	79.5	6.0	346	1	GCP_BACSU	O05518 bacillus su
33	79	5.9	304	1	HEM3_AQUAE	O66621 aquifex aeo

ALIGNMENTS

RESULT 1
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D26185; BAA05305.1; --
EMBL; Z99104; CAB11846.1; --
PIR; S66100; S66100.
Subtilist; BG10133; yacB.
InterPro; IPR004619; Baf.
Pfam; PF03309; Bvg_acc_factor; 1.
TIGRFAMS; TIGR00671; baf; 1.
Hypothetical protein; Complete proteome.
SEQUENCE 233 AA; 26217 MW; AAES96E732C15DF44 CRC64;

Query Match 83.4%; Score 1114; DB 1; Length 233;
Best Local Similarity 99.5%; Pred. No. 2e-85;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLVIDVGNNTVTLGVYHDKLEYHWRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60
Db :|||||
1 MLLVIDVGNNTVTLGVYHDKLEYHWRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60

QY 61 IISVVPPIMFALERMCCKYFHIETQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120
Db |||||
61 IISVVPPIMFALERMCCKYFHIETQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIHL 120

QY 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNI 180
Db |||||
121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNI 180

QY 181 GKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQD 214
Db |||||
181 GKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQD 214

RESULT 2
BAF BORPE
ID BAF BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP504;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshazer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required for transcription of the pertussis toxin operon in Escherichia coli."
RT J. Bacteriol. 177:3801-3807(1995).
RN [2]
RP SEQUENCE OF 1-38 FROM N.A.
RC STRAIN=BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a birA homolog in Bordetella pertussis."
RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 239-267 FROM N.A.
RC STRAIN=BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;

RT "The identification, cloning and mutagenesis of a genetic locus required for lipopolysaccharide biosynthesis in Bordetella pertussis."
RL Mol. Microbiol. 19:37-52(1996).
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF RNA POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12020; AAA75361.1; --
DR EMBL; AF016461; AAC68834.1; --
DR EMBL; X90711; CAA62242.1; --
DR PIR; I40327; I40327.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Transcription regulation; Activator.
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

Query Match 91.7%; Score 129.5; DB 1; Length 267;
Best Local Similarity 31.0%; Pred. No. 0.0011;
Matches 49; Conservative 22; Mismatches 68; Indels 19; Gaps 6;

QY 83 IEPQIVGPGMKTGLNPKYDNPKEVGADR---IVNAVAIHLNPLIVDFGTATTCYI 139
Db :|||
83 LRAQPLAMGLRNG---YRNPDLGADRWACWGVGLARQPSVHPPLLVASFGTATLDTI 138

QY 140 DENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNIIGKNTVSAMQSGILFGYVQ 199
Db :|||
139 GPDNVFPFGLILPGPAMMRGALAYGTAHLPLADGLVADYPI--DTHQAIASGIA---AAQ 193

QY 200 VEGIVKRMKWQA-----KQDLKVIATGGLAPLIANESD 232
Db |||
194 AGAIVR--QWLAGRQRYGOAPEIYVAGGWPEVRQEA 229

RESULT 3
GRSB BACBR
ID GRSB BACBR STANDARD; PRT; 4451 AA.
AC P14688;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Gramicidin S synthetase II [includes: ATP-dependent proline adenylase (ProA) (Proline activase); ATP-dependent valine adenylase (ValA) (Valine activase); ATP-dependent ornithine adenylase (OrnA) (Ornithine activase); ATP-dependent leucine adenylase (LeuA) (Leucine activase)].
GN GRSB OR GRS2.
OS Bacillus brevis (Brevibacillus brevis).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 9999 /IFO 15520 /JCM 8504 /NCTC 7096;
RX MEDLINE=92219998; PubMed=1560782;
RA Turgay K., Krause M., Marahiel M.A.;
RT "Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes."
RL Mol. Microbiol. 6:529-546(1992).
RN [2]
RP SEQUENCE OF 1-948 FROM N.A.
RC STRAIN=Nagano;
RX MEDLINE=92041751; PubMed=1939016;
RA Hori K., Yamamoto Y., Tokita K., Saito F., Kurotsu T., Kanda M., Okamura K., Furuyama J., Saito Y.;
RT "The nucleotide sequence for a proline-activating domain of

RT gramicidin S synthetase 2 gene from Bacillus brevis.";

RL J. Biochem. 110:111-119(1991).

RN [3]

RP SEQUENCE OF 1-143 FROM N.A.

RC STRAIN=ATCC 9999 / IF0 15520 / JCM 8504 / NCTC 7096;

RX MEDLINE=90008776; PubMed=2477357;

RA Kraetzschmar J., Krause M., Marahiel M.A.;

RT "Gramicidin S biosynthesis operon containing the structural genes

RT grsA and grsB has an open reading frame encoding a protein homologous

RT to fatty acid thioesterases.";

RL J. Bacteriol. 171:5422-5429(1989).

RN [4]

RP SEQUENCE OF 1-15, AND CHARACTERIZATION.

RC STRAIN=Nagano;

RX MEDLINE=92011463; PubMed=1917901;

RA Kurotsu T., Hori K., Kanda M., Saito Y.;

RT "Characterization and location of the L-proline activating fragment

RT from the multifunctional gramicidin S synthetase 2.";

RL J. Biochem. 109:763-769(1991).

CC -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME, ABLE TO

CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS PRO; VAL, ORN AND LEU.

CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.

CC -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.

CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE

CC ANTIOTIC GRAMICIDIN S (D-PHE-PRO-VAL-ORN-LEU)2.

CC -!- SUBUNIT: LARGE MULTIENZYME COMPLEX OF GRSA AND GRSB.

CC -!- DOMAIN: CONSISTS OF FOUR MODULES, AND HARBORS A PUTATIVE

CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE

CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE

CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE

CC ADENYLATION, THIOLATION, CONDENSATION (NOT FOR THE INITIATION

CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION

CC (OPTIONAL).

CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME

CC FAMILY..

CC -!- SIMILARITY: Contains 4 acyl carrier domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL; X61658; CAA43838.1; -.

DR EMBL; M29703; AAA58719.1; -.

DR EMBL; X15577; CAA33604.1; -.

DR EMBL; D00938; BAA00778.1; -.

DR PIR; S20542; YGBSG2.

DR HSSP; P14687; 1AMU.

DR InterPro; IPR000873; AMP-bind.

DR InterPro; IPR001242; Condensatn.

DR InterPro; IPR006163; Pp_bind.

DR InterPro; IPR006162; Ppantne attach.

DR InterPro; IPR000379; Ser_estrs_site.

DR InterPro; IPR001031; Thioesterase.

DR Pfam; PF00501; AMP-binding; 4.

DR Pfam; PF00668; Condensation; 4.

DR Pfam; PF00550; pp-binding; 4.

DR Pfam; PF00975; Thioesterase; 1.

DR PROSITE; PS00012; THIOESTERASE; 4.

DR PROSITE; PS00455; AMP_BINDING; 4.

DR PROSITE; PS00075; ACP_DOMAIN; 4.

KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;

KW Repeat; Phosphopantetheine; Hydrolase.

FT INIT_MET 0

FT REPEAT 466 1043 DOMAIN 1 (PROLINE-ACTIVATING).

FT REPEAT 1521 2080 DOMAIN 2 (VALINE-ACTIVATING).

FT REPEAT 2538 3135 DOMAIN 3 (ORNITHINE-ACTIVATING).

FT REPEAT 3591 4173 DOMAIN 4 (LEUCINE-ACTIVATING).

FT DOMAIN 975 1042 ACYL CARRIER (ACP) 1.

FT DOMAIN 2011 2078 ACYL CARRIER (ACP) 2.

FT DOMAIN 3057 3124 ACYL CARRIER (ACP) 3.

FT DOMAIN 4095 4162 ACYL CARRIER (ACP) 4.

FT BINDING 1005 1005 PHOSPHOPANTETHEINE (BY SIMILARITY).

FT BINDING 2041 2041 PHOSPHOPANTETHEINE (BY SIMILARITY).

FT BINDING 3087 3087 PHOSPHOPANTETHEINE (BY SIMILARITY).

FT BINDING 4125 4125 PHOSPHOPANTETHEINE (BY SIMILARITY).

FT CONFLICT 274 274 H -> D (IN REF. 2).

FT CONFLICT 418 418 K -> R (IN REF. 2).

FT CONFLICT 654 664 NAVLTCTVTKKF -> TCSFDVCYQEI (IN REF. 2).

FT CONFLICT 941 946 HVRLHL -> QLPLTP (IN REF. 2).

SQ SEQUENCE 4451 AA; 510036 MW; E0029C9B51F5A4B7 CRC64;

Query Match 6.8%; Score 90.5; DB 1; Length 4451;

Best Local Similarity 22.7%; Pred. No. 48;

Matches 56; Conservative 35; Mismatches 77; Indels 79; Gaps 14;

QY 70 MFALERMCTKYFHIEPQIVGPGMKTGLNKKYDNPKEVG-----ADRIV 112

Db 1396 MFVLQNTDRKSEFEVQITITPYVPNSRHSKFDLTLEVSEEQNEILLCLEYCTKLFTDKTV 1455

QY 113 NAVAA-----IH-LYGNPLIV-----DFG-TATTYCYIDENKQYM--- 146

Db 1456 ERMAGHFLQILHAIVGNPTIIIEIILSEEEKQHILFEFNDTKTTPYPLCKQPKDYLRNR 1515

QY 147 ----GGAIAPG---ITISTEALYSRAAKLPRI---EITRPDNIIGKNTVSAMQSGILFGY 196

Db 1516 WRRRADHVAVGWKDQTLTYRELNERANQAVRLRQKGVQPDNIVG--LLVRSPEMLVG- 1572

QY 197 VGQVEGIVK-----RMKWQAKQDLKV---IATGGLAPLIANESDCIDIVD 238

Db 1573 --IMGILKAGGAYLPDPEYPADRIISYMI-QDCGVRIMLTQQHLLSLVHDEFDCV-ILD 1627

QY 239 PFLTLLKG 245

Db 1628 EDSLYKG 1634

RESULT 4

BUK_THETN STANDARD; PRT; 357 AA.

ID BUK_THETN

AC Q8R832;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable butyrate kinase (EC 2.7.2.7) (BK) (Branched-chain carboxylic acid kinase).

GN BUK OR TTE2201.

OS Thermoanaerobacter tengcongensis.

OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;

OC Thermoanaerobacteriaceae; Thermoanaerobacter.

OX NCBI_TaxID=119072;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MB4 / JCM 11007;

RX MEDLINE=21992816; PubMed=11997336;

RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,

RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,

RA Tan H., Chen R., Wang J., Yu J., Yang H.;

RT "A complete sequence of T. tengcongensis genome.";

RL Genome Res. 12:689-700(2002).

CC -!- CATALYTIC ACTIVITY: ATP + 2-butanolate = ADP + butanoyl phosphate.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the acetokinase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL; AE013165; AAM25356.1; -.

```
DR HAMAP; MF 00542; -; 1.
DR InterPro; IPR000890; Acetate_kin.
DR Pfam; PF00871; Acetate kinase; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
DR PROSITE; PS01076; ACETATE_KINASE_2; 1.
KW Transferase; Kinase; Complete proteome.
SQ SEQUENCE 357 AA; 39217 MW; 05CDD09A2E7EF3F CRC64;

Query Match          6.7%; Score 89.5; DB 1; Length 357;
Best Local Similarity 19.9%; Pred. No. 3.2;
Matches 66; Conservative 43; Mismatches 94; Indels 129; Gaps 16;

QY 1 LLLVIDGNTNTVLGVYHDGKLEYHWRIETSRHKT-----DEFGMILRSLFD--- 48
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 5 LILVINPGSTSTKAVFRDKEPVF---TETLRHSTEELSKYKRIIDQFEFRTQAILMLK 61
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 49 HSGLMFEQIDGII-----ISS----- 64
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 62 EKGISLSQIDAIVGRGGLLKPIESGTYIVNEKMLEDLKKAERGEHASNLGAIAYTLAKE 121
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 65 -----VPPIMPALERMC-----TKYFH-IEPQIVGPGMKTGLNIKYDNPKEV 106
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 122 HNIPAYIVDPVVVDELEDVARITGLPRTEKQSFHALNQKAIARRLASDLGKRYDE---- 177
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 107 GADRIVNAVAAILHLYGNPLIVDFGTATTTCYIDENKQYMG-GAIAPGITISTEALYSRA 165
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 178 -----VNLIIA-HLGG-----ISVGAHRKGRVIDVNDALNGEGFPSP-----ERA 217
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 166 AKLPRIETRPDNIIGKNTVSAMQ-----SGILFGYVG-----QVEGIV 204
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 218 GGLPVLDLVKL-CYSGKYTFEEMKKLIGKGGIVAHGLTNDVREVYKMIENGDKNAELIL 276
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 205 KRMKWQAKQDLKVIATGGLAPLIANESDCIDI 236
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 277 DAMAYQTAKET-----GSMVAVLKGKVDAGI 303
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
RESULT 5
PURA LEPIN
ID PURA LEPIN STANDARD; PRT; 425 AA.
AC Q8F738;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)
DE (Adss) (AMPSase).
GN PURA OR LA1110.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX MEDLINE=22598143; PubMed=12712204;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.;
RT "Unique physiological and pathogenic features of Leptospira
interrogans revealed by whole-genome sequencing.";
RL Nature 422:888-893(2003).
CC -!- FUNCTION: Plays an important role in the de novo pathway of purine
nucleotide biosynthesis.
CC -!- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +
adenylosuccinate.
CC -!- PATHWAY: AMP biosynthesis; first committed step.
CC -!- SIMILARITY: Belongs to the adenylosuccinate synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE011293; AAN48309.1; -.
DR HAMAP; MF 00011; -; 1.
DR InterPro; IPR001114; Asucc_synthase.
DR Pfam; PF00709; Adenylsucc synt; 1.
DR ProDom; PD001188; Asucc_synthase; 1.
DR TIGRFAMs; TIGR00184; purA; 1.
DR PROSITE; PS01266; ADENYLOSUCCIN SYN 1; FALSE_NEG.
DR PROSITE; PS00513; ADENYLOSUCCIN SYN 2; FALSE_NEG.
KW Purine biosynthesis; Ligase; GTP-binding; Complete proteome.
FT ACT_SITE 141 141 BY SIMILARITY.
FT ACT_SITE 148 148 BY SIMILARITY.
SQ SEQUENCE 425 AA; 47057 MW; 8B57C6DF6FB9CF95 CRC64;

Query Match          6.6%; Score 88.5; DB 1; Length 425;
Best Local Similarity 20.6%; Pred. No. 4.8;
Matches 59; Conservative 40; Mismatches 104; Indels 83; Gaps 15;

QY 7 VGNNTNTVLGV-YHDGKLEYHWRI-----ETSRHKTDEFGMILRSLFDHSGLMFEQIDGII 61
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 127 IGTTKKGIGICYADKMMRTGLRVGDLLDTS-YQTR-----LKHLDVEKNRELDKLYG-- 177
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 62 ISSVVPPIMPALERMCTKYF--HIEPQIVGPG-----MKTGLNIKYDNPKEVGADRIVN 113
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 178 ---MPPVSYNDINEGLKFFLSKVKNIIINTAYYLDTELKKGKRVLLLEGAQGTGLD---- 229
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 114 AVAAIHLNGNPLIVDFGTATTTCYIDENKQYMGGA-IAPGITIS-----TEALYSR 164
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 230 -----VDFG---TPYVVTSSNPTTGGALIGTGIPFQHLKHVIGITKAYTTR 272
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 165 AAKLPRIETRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLK-VIATGGL 223
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 273 VGEGP-----FPTTELLGEAGEKLQKG---GEFGATTGPRRCGWFVDEMLKHSVRINGI 324
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 224 APLIANESDCIDIVD-----PFLTLKGLELIYE 251
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 325 TSIALTAKIDILSDYDTIPVATGYKLNKLTLCFSPSQGLDKVEVIYE 370
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
RESULT 6
HI61_PSEAE
ID HI61_PSEAE STANDARD; PRT; 256 AA.
AC Q9HU44;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF1 (EC 4.1.3.-) (IGP
synthase cyclase subunit) (IGP synthase subunit hisF1) (ImGP synthase
subunit hisF1) (IGPS subunit hisF1).
GN HISF1 OR PA5140.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
```

CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-
CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC
CC -!- PATHWAY: Histidine biosynthesis; fifth step.
CC -!- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004927; AAC08525.1; -.
CC PIR; H83002; H83002.
CC HAMAP; MF_01013; -; 1.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006062; His_biosynth.
CC InterPro; IPR004651; HisF.
CC Pfam; PF00977; His_biosynth; 1.
CC TIGRFAMs; TIGR00735; hisF; 1.
KW Histidine biosynthesis; Lyase; Complete proteome.
FT ACT_SITE 12 12 POTENTIAL.
FT ACT_SITE 131 131 POTENTIAL.
SQ SEQUENCE 256 AA; 27131 MW; 6486A76CD308FAA7 CRC64;

Query Match 6.6%; Score 88; DB 1; Length 256;
Best Local Similarity 22.7%; Pred. No. 3;
Matches 60; Conservative 45; Mismatches 79; Indels 80; Gaps 15;
QY 5 IDVGNTNTVLGVYHDKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIISS 64
Db 11 LDVDNGRVWKGKVFENIRDAGDPVEIARR-----YDEQGA--DEITFLDITA 55
QY 65 VV---PPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVAIAHLY 121
Db .56 SVDGRDITLHTVERMASQVF--IPLTVGGVRSVQDIR--NLLNAGADKVSINTAAVF-- 109
QY 122 GNPLIVD---FGTATTCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIE---- 173
Db 110 -NPEFVGEADRFGSQICIVVAIDAKK-----VSAPG-----EAPRWEIFTHG 150
QY 174 ----TRPDNII-----GKNTVSAM-QSGILFGY-VGQVEGIVKRMKWQAKDLKV 217
Db 151 GRKPTGLDAVLWAKMEDLGAGEILLTSMDDQGVKSGYDLGVTRAISEAV-----NVPV 204
QY 218 IATGGL-----APLIANESDCI 234
Db 205 IASGGVGNLEHIAAGILEGKADAV 228

RESULT 7
SAHH_METKA STANDARD; PRT; 424 AA.
ID SAHH_METKA
AC P58855;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adenosylhomocysteinease (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHCY OR MK0368.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -!- COFACTOR: NAD (By similarity).
CC -!- PATHWAY: Activated methyl cycle.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the adenosylhomocysteinease family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010334; AAM01583.1; -.
CC HAMAP; MF_00563; -; 1.
CC InterPro; IPR000043; Ado_hcyase.
CC Pfam; PF05221; AdoHcyase; 1.
CC Pfam; PF00670; AdoHcyase_NAD; 1.
CC TIGRFAMs; TIGR00936; ahcy; 1.
CC PROSITE; PS00738; ADOHCYASE_1; FALSE_NEG.
CC PROSITE; PS00739; ADOHCYASE_2; 1.
KW Hydrolase; NAD; One-carbon metabolism; Complete proteome.
FT NP_BIND 213 244 NAD (POTENTIAL).
SQ SEQUENCE 424 AA; 47015 MW; 3DE7B10A3B44E1D5 CRC64;

Query Match 6.6%; Score 87.5; DB 1; Length 424;
Best Local Similarity 20.8%; Pred. No. 5.8;
Matches 62; Conservative 47; Mismatches 96; Indels 93; Gaps 14;
QY 19 DGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIISSVVPPIMF-----AL 73
Db 97 EGVHVIYAWRGE-----TEEY-----YQNI DRV L--SHEPDIIVDDGADCI 135
QY 74 ERMCTKYFHIEPQIVGPGMK--TGLN-----IKY--DNPKEVG 107
Db 136 ARVHTEFPDLAERVIGATEETTTGVNRLHAMHREGVLKFPVIAVNDAKTKYLDNRYGTG 195
QY 108 ADRIVNAVA--IHLVGNPLIVDFGTATTCYIDENKQYMG-GAIAPGITI----STEA 160
Db 196 QSALDGLMRATNILLAGKTVVVVGYG---WCGRGIARRARGLGANVIVVEVDPIKAMEA 251
QY 161 LY-----SRAAKLPRIETRPDN--IIGKNTVSAMQSGILFGYVGQVE----- 201
Db 252 IFDGFVRVMPMDRAAEEDIFITATGNRDVIRGEHIEKMKDGVILANAGHFDVEIDKEYLE 311
QY 202 -----GIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPFLTKGLEL 248
Db 312 EHCEEKIDRRGGLVTEYRMPDGGKRVYLLIAEGRVLNLAAGEGHPHIEIMDISFALQALSV 369

RESULT 8
PYRD_SALTI STANDARD; PRT; 336 AA.
ID PYRD_SALTI
AC Q82759;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase)
DE (DHODase) (DHOD).
GN PYRD OR STY1079 OR T1862.
OS Salmonella typhi.

Best Local Similarity 25.0%; Pred. No. 7.2;
Matches 51; Conservative 27; Mismatches 70; Indels 56; Gaps 11;
QY 95 GUNI--KYDNPKEVGADRIVNAVAIAHLYGNPLIV-----VDFGTATTVCYID 140
Db 137 GINIGKNKDTPEQKDDYLICMEKIYAYAGYIAINISSPNTPLRTLOQGEA-----LD 191
QY 141 E-----NKQ-----YMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNIIGKNT- 184
Db 192 DLLTAIKNQNDLQAMHKKYVPIAVKIAPDLSEELIQVADSLVRHNI---DGVIAITNTT 248
QY 185 -----VSAM-----QSGILFGYGVQVEG--IVKRMKWAQDLKVIATGGLAPLANE--- 230
Db 249 LDRSLVQGMKNCDQTGGSLGRPLQLKSTEIIRRLSLELNGRLPIIGVGIDSIVIAAREKI 308
QY 231 ---SDCIDIVDPFLTLKGLLEIYE 251
Db 309 AAGASLVQIYSGFI-FKGPPLIKE 331

RESULT 13
DMSB_ECOLI STANDARD; PRT; 204 AA.
ID DMSB_ECOLI STANDARD; PRT; 204 AA.
AC P18776; P77745;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Anaerobic dimethyl sulfoxide reductase chain B (DMSO reductase iron-
DE sulfur subunit).
GN DMSB OR B0895.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-9.
RC STRAIN=K12 / C600;
RX MEDLINE=89096500; PubMed=3062312;
RA Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H.;
RT "Nucleotide sequence of the dmsABC operon encoding the anaerobic
RT dimethylsulphoxide reductase of Escherichia coli.";
RL Mol. Microbiol. 2:785-795(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kaji-hara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP EPR SPECTROSCOPY OF IRON-SULFUR CLUSTERS.
RX MEDLINE=91070067; PubMed=2174699;
RA Cammack R., Weiner J.H.;
RT "Electron paramagnetic resonance spectroscopic characterization of
RT dimethyl sulfoxide reductase of Escherichia coli.";
RL Biochemistry 29:8410-8416(1990).
RN [5]

MUTAGENESIS.
MEDLINE=91355180; PubMed=1653010;
RX Rothery R.A., Weiner J.H.;
RT "Alteration of the iron-sulfur cluster composition of Escherichia
RT coli dimethyl sulfoxide reductase by site-directed mutagenesis.";
RL Biochemistry 30:8296-8305(1991).
CC -!- FUNCTION: Electron transfer subunit of the terminal reductase
CC during anaerobic growth on various sulfoxide and N-oxide
CC compounds.
CC -!- COFACTOR: Binds 4 4Fe-4S clusters.
CC -!- SUBUNIT: Heterotrimeric enzyme composed of a catalytic heterodimer
CC (DmsAB) and a membrane anchor protein (DmsC).
CC -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03412; AAA83844.1; -.
DR EMBL; AE000191; AAC73981.1; -.
DR EMBL; D90727; BAA35627.1; -.
DR PIR; F64828; F64828.
DR HSSP; P00195; 1CLF.
DR EcoGene; EG10233; dmsB.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR Pfam; PF00037; fer4; 1.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW Electron transport; 4Fe-4S; Iron-sulfur; Complete proteome.
FT INIT MET 0 0
FT METAL 13 13 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 16 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 19 19 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 23 23 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 66 66 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 69 69 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 74 74 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 78 78 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 98 98 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 101 101 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 104 104 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 108 108 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 125 125 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 128 128 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 140 140 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 144 144 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT MUTAGEN 101 101 C->F,S,W,Y: LOSS OF ELECTRON TRANSFER
FT FROM MENAQUINOL TO DMSO.
FT CONFLICT 169 169 P -> PRA (IN REF. 1).
SQ SEQUENCE 204 AA; 22738 MW; 90425D307DEF2130 CRC64;
Query Match 6.3%; Score 83.5; DB 1; Length 204;
Best Local Similarity 23.3%; Pred. No. 5.4;
Matches 45; Conservative 27; Mismatches 66; Indels 55; Gaps 10;
QY 15 GYVHDGKLEYHRIETSRHKTEDEFGMILRSLFDHSGMLPFIQDIISSVVPIMFALE 74
Db 51 GVVHQNVFAYY--LSISCNHCEDP---ACTKVCPSGAMHKREDGFVVVD-----E 95
QY 75 RMC--TKYFHIEPQIVGPGMKTGLNIKYDNPK-----EVGADRIVNAVAIAHLYGNPLI 126
Db 96 DVCIGCRYCHMACPYGAP-----QYNETHKGMTKDCGYDRVAEGKKPICVESCPLR 147
QY 127 VVDFGTATTVCYIDENKQYMG--GAIAPGITISTEALYSRAAKLPRIETRPDNIIGKNT 184
Db 148 ALDFGP-----IDELRKKHGDIAAVAP-----LPRHFTKP-NIVIKPN 185
QY 185 VSAMQSGILFGYV 197


```
Db      186 ANSRPTGDTGYL 198

RESULT 14
ATKC_RHILO
ID _ATKC_RHILO      STANDARD;      PRT;      187 AA.
AC  Q98GX7;
DT  15-SEP-2003 (Rel. 42, Created)
DT  15-SEP-2003 (Rel. 42, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  Potassium-transporting ATPase C chain (EC 3.6.3.12) (Potassium-
DE  translocating ATPase C chain) (ATP phosphohydrolase [potassium-
DE  transporting] C chain) (Potassium binding and translocating subunit
DE  C).
GN  KDPC OR MLL3129.
OS  Rhizobium loti (Mesorhizobium loti).
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC  Phyllobacteriaceae; Mesorhizobium.
OX  NCBI_TaxID=381;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MAFF303099;
RX  MEDLINE=21082930; PubMed=11214968;
RA  Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA  Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA  Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA  Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA  Takeuchi C., Yamada M., Tabata S.;
RT  "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT  Mesorhizobium loti.";
RL  DNA Res. 7:331-338(2000).
CC  -!- FUNCTION: One of the components of the high-affinity ATP-driven
CC  potassium transport (or KDP) system, which catalyzes the
CC  hydrolysis of ATP coupled with the exchange of hydrogen and
CC  potassium ions. The C subunit may be involved in assembly of the
CC  KDP complex (By similarity).
CC  -!- CATALYTIC ACTIVITY: ATP + H(2)O + K(+) (Out) = ADP + phosphate +
CC  K(+) (In).
CC  -!- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC  (Probable).
CC  -!- SIMILARITY: Belongs to the kdpC family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AP003001; BAB50089.1; -.
CC  HAMAP; MF_00276; -. 1.
CC  InterPro; IPR003820; K_ATPase_KdpC.
CC  Pfam; PF02669; KdpC; 1.
CC  TIGRFAMs; TIGR00681; kdpC; 1.
KW  Hydrolase; Transport; Potassium transport; Transmembrane;
KW  Inner membrane; Complete proteome.
FT  TRANSMEM 9 31 Potential.
SQ  SEQUENCE 187 AA; 19572 MW; 47E91214E671BCB2 CRC64;

Query Match      6.2%; Score 83; DB 1; Length 187;
Best Local Similarity 27.1%; Pred. No. 5.4;
Matches 46; Conservative 20; Mismatches 64; Indels 40; Gaps 10;

QY  42 ILRLSPDH--SGLMFEQIDGIIISVVPPIMFALERMCTKYEHIEPQIVPGMKTGL---- 96
| :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| :
Db  30 IAQALFPRQANGSLIEK-DGKVGISELIGQAFASD----KYFHCGRPSAAGNGYDAGASGG 84
| :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| :

QY  97 -NIKYDNPKEV-----GADRIVNAVAATHLYGNPLIVVDFGTATTTCYIDENKQYMGGA 150
| :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| :
Db  85 SNLGPTNPKLIERIKGGAEKLAEE-----NPNQVPVMDLVTT-----SGSGL 126
| :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| :

QY  151 APGITISTEALYSRAAKLPRIETRP-DNIIGKNTVSAMQSGILFGYVGQ 199
```

```
Db      127 DP--QISPEAAY---FQVPRVAKARGIDEAKVKALVDGQVGEGLGFMGE 171

RESULT 15
PIMT_VIBPA
ID _PIMT_VIBPA      STANDARD;      PRT;      208 AA.
AC  Q87LQ6;
DT  15-SEP-2003 (Rel. 42, Created)
DT  15-SEP-2003 (Rel. 42, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  Protein-L-isoaspartate O-methyltransferase (EC 2.1.1.77) (Protein-
DE  beta-aspartate methyltransferase) (PIMT) (Protein L-isoaspartyl
DE  methyltransferase) (L-isoaspartyl protein carboxyl methyltransferase).
GN  PCM OR VP2555.
OS  Vibrio parahaemolyticus.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC  Vibrionaceae; Vibrio.
OX  NCBI_TaxID=670;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=RIMD 2210633 / Serotype O3:K6;
RX  MEDLINE=22508454; PubMed=12620739;
RA  Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA  Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA  Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT  "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT  distinct from that of V. cholerae.";
RL  Lancet 361:743-749(2003).
CC  -!- FUNCTION: Catalyzes the methyl esterification of L-isoaspartyl
CC  residues in peptides and proteins that result from spontaneous
CC  decomposition of normal L-aspartyl and L-asparaginyl residues. It
CC  plays a role in the repair and/or degradation of damaged proteins
CC  (By similarity).
CC  -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + protein L-beta-
CC  aspartate = S-adenosyl-L-homocysteine + protein L-beta-aspartate
CC  methyl ester.
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC  -!- SIMILARITY: BELONGS TO THE L-ISOASPARTYL/D-ASPARTYL PROTEIN
CC  METHYLTRANSFERASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AP005081; BAC60818.1; -.
CC  HAMAP; MF_00090; -. 1.
CC  PROSITE; PS01279; PCMT; 1.
KW  Transferase; Methyltransferase; Complete proteome.
SQ  SEQUENCE 208 AA; 23041 MW; C3B0E83DAB4F9CFE CRC64;

Query Match      6.2%; Score 83; DB 1; Length 208;
Best Local Similarity 31.9%; Pred. No. 6;
Matches 22; Conservative 15; Mismatches 28; Indels 4; Gaps 2;

QY  150 IAPGITISTEALYSRAAKLPRIETRPDNIIGKNTVSAMQSGILFGYVGQVEGI--VKRM 207
| :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| :
Db  52 IGGQTISQPIYIVARMTTEL--LELQRASNVLIEIGTSGYQTAVLAQIVDHVYSVERIKSL 109
| :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| :

QY  208 KWQAKQDLK 216
| :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| :
Db  110 QWEAKRRLLK 118
| :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| : :||| :
```

Search completed: December 18, 2003, 13:02:08
Job time : 19 secs